THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 92, Number 1 January 1985

## Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, University of California, Berkeley, CA 94720. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS CO., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.




The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# THE AMERICAN

# MATHEMATICAL MONTHLY

AN OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

DEVOTED TO THE INTERESTS OF COLLEGIATE MATHEMATICS



THE AMERICAN MATHEMATICAL MONTHLY, FOUNDED IN 1894 BY BENJAMIN F. FINKEL, WAS PUBLISHED BY HIM UNTIL 1913. FROM 1913 TO 1916 IT WAS OWNED AND PUBLISHED BY REPRESENTATIVES OF FOURTEEN UNIVERSITIES AND COLLEGES IN THE MIDDLE WEST

VOLUME 92

1985

PUBLISHED BY THE ASSOCIATION
MONTPELIER, VT, AND WASHINGTON, DC

# THE $3x + 1$ PROBLEM AND ITS GENERALIZATIONS

JEFFREY C. LAGARIAS
*AT & T Bell Laboratories, Murray Hill, NJ 07974*

**1. Introduction.** The $3x + 1$ *problem*, also known as *the Collatz problem*, *the Syracuse problem*, *Kakutani's problem*, *Hasse's algorithm*, and *Ulam's problem*, concerns the behavior of the iterates of the function which takes odd integers $n$ to $3n + 1$ and even integers $n$ to $n/2$. The $3x + 1$ Conjecture asserts that, starting from any *positive* integer $n$, repeated iteration of this function eventually produces the value 1.

The $3x + 1$ Conjecture is simple to state and *apparently* intractably hard to solve. It shares these properties with other iteration problems, for example that of aliquot sequences (see Guy [**36**], Problem B6) and with celebrated Diophantine equations such as Fermat's last theorem. Paul Erdős commented concerning the intractability of the $3x + 1$ problem: "Mathematics is not yet ready for such problems." Despite this doleful pronouncement, study of the $3x + 1$ problem has not been without reward. It has interesting connections with the Diophantine approximation of $\log_2 3$ and the distribution (mod 1) of the sequence $\{(3/2)^k\colon\ k = 1, 2, \dots\}$, with questions of ergodic theory on the 2-adic integers $\mathbf{Z}_2$, and with computability theory—a generalization of the $3x + 1$ problem has been shown to be a computationally unsolvable problem. In this paper I describe the history of the $3x + 1$ problem and survey all the literature I am aware of about this problem and its generalizations.

The exact origin of the $3x + 1$ problem is obscure. It has circulated by word of mouth in the mathematical community for many years. The problem is traditionally credited to Lothar Collatz, at the University of Hamburg. In his student days in the 1930's, stimulated by the lectures of Edmund Landau, Oskar Perron, and Issai Schur, he became interested in number-theoretic functions. His interest in graph theory led him to the idea of representing such number-theoretic functions as directed graphs, and questions about the structure of such graphs are tied to the behavior of iterates of such functions [**25**]. In his notebook dated July 1, 1932, he considered the function

$$g(n) = \begin{cases} \dfrac{2}{3}n, & \text{if } n \equiv 0 \pmod 3, \\[2mm] \dfrac{4}{3}n - \dfrac{1}{3}, & \text{if } n \equiv 1 \pmod 3, \\[2mm] \dfrac{4}{3}n + \dfrac{1}{3}, & \text{if } n \equiv 2 \pmod 3, \end{cases}$$

which gives rise to a permutation $P$ of the natural numbers

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \\ 1 & 3 & 2 & 5 & 7 & 4 & 9 & 11 & 6 & \dots \end{pmatrix}.$$

He posed the problem of determining the cycle structure of $P$, and asked in particular whether or not the cycle of this permutation containing 8 is finite or infinite, i.e., whether or not the iterates $g^{(k)}(8)$ remain bounded or are unbounded [**24**]. I will call the study of the iterates of $g(n)$ the

*Jeffrey C. Lagarias*: I was first exposed to the $3x + 1$ problem in 1967 as a high school student working at the National Bureau of Standards. Afterwards I worked on it from time to time. Out of curiosity and frustration I gradually became a historian of the problem, accumulating a collection of papers about it. This survey is a happy consequence. I obtained a Ph.D. (1974) in analytic number theory at M.I.T. under the supervision of Harold Stark. I have been on the staff of AT & T Bell Laboratories since then, and have held visiting positions at the University of Maryland (mathematics) and Rutgers University (computer science). My research interests include computational complexity theory, number theory, and cryptography.

*original Collatz problem*. Although Collatz never published any of his iteration problems, he circulated them at the International Congress of Mathematicians in 1950 in Cambridge, Massachusetts, and eventually the original Collatz problem appeared in print ([**9**], [**47**], [**59**]). His original question concerning $g^{(k)}$ (8) has never been answered; the cycle it belongs to is believed to be infinite. Whatever its exact origins, the $3x + 1$ problem was certainly known to the mathematical community by the early 1950's; it was discovered in 1952 by B. Thwaites [**69**].

During its travels the $3x + 1$ problem has been christened with a variety of names. Collatz's colleague H. Hasse was interested in the $3x + 1$ problem and discussed generalizations of it with many people, leading to the name *Hasse's algorithm* [**40**]. The name *Syracuse problem* was proposed by Hasse during a visit to Syracuse University in the 1950's. Around 1960, S. Kakutani heard the problem, became interested in it, and circulated it to a number of people. He said "For about a month everybody at Yale worked on it, with no result. A similar phenomenon happened when I mentioned it at the University of Chicago. A joke was made that this problem was part of a conspiracy to slow down mathematical research in the U.S. [**45**]." In this process it acquired the name *Kakutani's problem*. S. Ulam also heard the problem and circulated the problem at Los Alamos and elsewhere, and it is called *Ulam's problem* in some circles ([**13**], [**69**]).

In the last ten years the $3x + 1$ problem has forsaken its underground existence by appearing in various forms as a problem in books and journals, sometimes without attribution as an unsolved problem. Prizes have been offered for its solution: \$50 by H. S. M. Coxeter in 1970, then \$500 by Paul Erdős, and more recently £1000 by B. Thwaites [**69**]. Over twenty research articles have appeared on the $3x + 1$ problem and related problems.

In what follows I first discuss what is known about the $3x + 1$ problem itself, and then discuss generalizations of the problem. I have included or sketched proofs of Theorems B, D, E, F, M and N because these results are either new or have not appeared in as sharp a form previously; the casual reader may skip these proofs.

**2. The $3x + 1$ problem.** The known results on the $3x + 1$ problem are most elegantly expressed in terms of iterations of the function

$$T(n) = \begin{cases} \dfrac{3n+1}{2}, & \text{if } n \equiv 1 \pmod 2, \\ \dfrac{n}{2}, & \text{if } n \equiv 0 \pmod 2. \end{cases} \tag{2.1}$$

One way to think of the $3x + 1$ problem involves a directed graph whose vertices are the positive integers and that has directed edges from $n$ to $T(n)$. I call this graph the *Collatz graph* of $T(n)$ in honor of L. Collatz [**25**]. A portion of the Collatz graph of $T(n)$ is pictured in Fig. 1. A directed graph is said to be *weakly connected* if it is connected when viewed as an undirected graph, i.e., for any two vertices there is a path of edges joining them, ignoring the directions on the edges. The $3x + 1$ Conjecture can be formulated in terms of the Collatz graph as follows.

$3x + 1$ CONJECTURE (First form). *The Collatz graph of $T(n)$ on the positive integers is weakly connected.*

We call the sequence of iterates $(n, T(n), T^{(2)}(n), T^{(3)}(n), \dots)$ the *trajectory* of $n$. There are three possible behaviors for such trajectories when $n > 0$.

(i) *Convergent trajectory*. Some $T^{(k)}(n) = 1$.
(ii) *Non-trivial cyclic trajectory*. The sequence $T^{(k)}(n)$ eventually becomes periodic and $T^{(k)}(n) \neq 1$ for any $k \geqslant 1$.
(iii) *Divergent trajectory*. $\lim_{k \to \infty} T^{(k)}(n) = \infty$.

The $3x + 1$ Conjecture asserts that all trajectories of positive $n$ are convergent. It is certainly true for $n > 1$ that $T^{(k)}(n) = 1$ cannot occur without some $T^{(k)}(n) < n$ occurring. Call the least

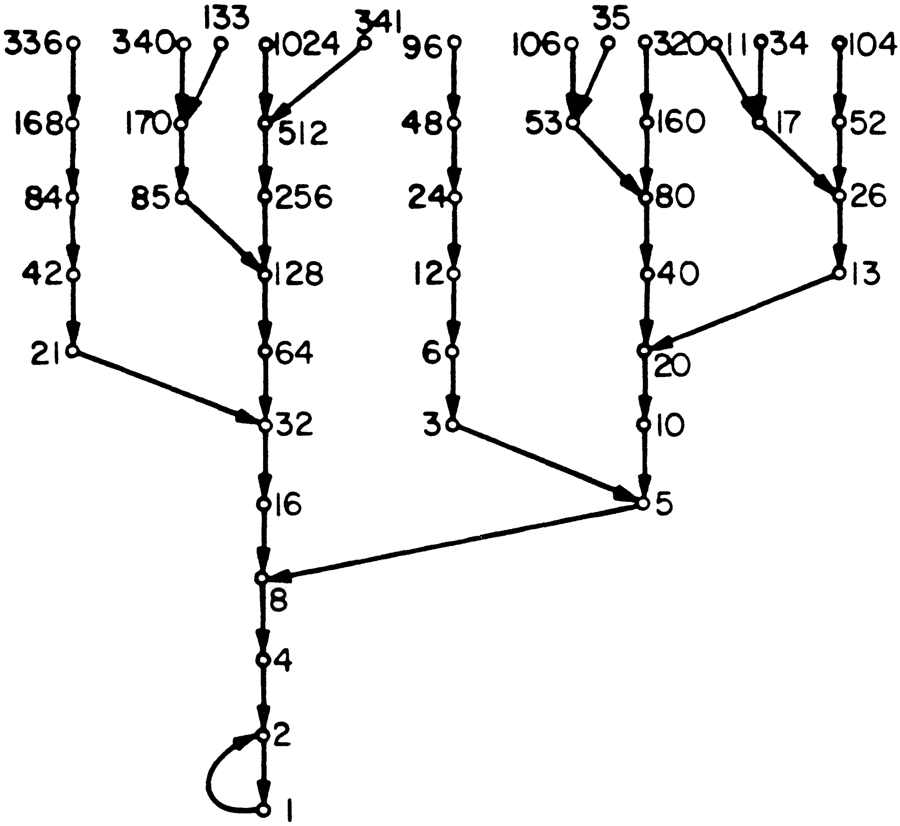



FIG. 1. A portion of the Collatz graph of iterates of the function $T(n)$.

positive $k$ for which $T^{(k)}(n) < n$ the *stopping time* $\sigma(n)$ of $n$, and set $\sigma(n) = \infty$ if no $k$ occurs with $T^{(k)}(n) < n$. Also call the least positive $k$ for which $T^{(k)}(n) = 1$ the *total stopping time* $\sigma_\infty(n)$ of $n$, and set $\sigma_\infty(n) = \infty$ if no such $k$ occurs. We may restate the $3x + 1$ Conjecture in terms of the stopping time as follows.

$3x + 1$ CONJECTURE (Second form). *Every integer $n \geqslant 2$ has a finite stopping time.*

The appeal of the $3x + 1$ problem lies in the irregular behavior of the successive iterates $T^{(k)}(n)$. One can measure this behavior using the stopping time, the total stopping time, and the *expansion factor* $s(n)$ defined by

$$s(n) = \frac{\sup\limits_{k \geqslant 0} T^{(k)}(n)}{n},$$

if $n$ has a bounded trajectory and $s(n) = +\infty$ if $n$ has a divergent trajectory. For example $n = 27$ requires 70 iterations to arrive at the value 1 and

$$s(27) = \frac{\sup\limits_{k \geqslant 0} T^{(k)}(27)}{27} = \frac{4616}{27} \approx 171.$$

Table 1 illustrates the concepts defined so far by giving data on the iterates $T^{(k)}(n)$ for selected values of $n$.

TABLE 1. Behavior of iterates $T^{(k)}(n)$.

| $n$ | $\sigma(n)$ | $\sigma_\infty(n)$ | $s(n)$ |
|---|---|---|---|
| 1 | $\infty$ | 2 | 2 |
| 7 | 7 | 11 | 3.7 |
| 27 | 59 | 70 | 171. |
| $2^{50}-1$ | 143 | 383 | $6.37 \times 10^8$ |
| $2^{50}$ | 1 | 50 | 1 |
| $2^{50}+1$ | 2 | 223 | 1.50 |
| $2^{500}-1$ | 1828 | 4331 | $1.11 \times 10^{88}$ |
| $2^{500}+1$ | 2 | 2204 | 1.50 |

The $3x + 1$ Conjecture has been numerically checked for a large range of values of $n$. It is an interesting problem to find efficient algorithms to test the conjecture on a computer. The current record for verifying the $3x + 1$ Conjecture seems to be held by Nabuo Yoneda at the University of Tokyo, who has reportedly checked it for all $n < 2^{40} \approx 1.2 \times 10^{12}$ [**2**]. In several places the statement appears that A. S. Fraenkel has checked that all $n < 2^{50}$ have a finite total stopping time; this statement is erroneous [**32**].

**2.1. A heuristic argument.** The following heuristic probabilistic argument supports the $3x + 1$ Conjecture (see [**28**]). Pick an odd integer $n_0$ at random and iterate the function $T$ until another odd integer $n_1$ occurs. Then $\frac{1}{2}$ of the time $n_1 = (3n_0 + 1)/2$, $\frac{1}{4}$ of the time $n_1 = (3n_0 + 1)/4$, $\frac{1}{8}$ of the time $n_1 = (3n_0 + 1)/8$, and so on. If one supposes that the function $T$ is sufficiently "mixing" that successive odd integers in the trajectory of $n$ behave as though they were drawn at random (mod $2^k$) from the set of odd integers (mod $2^k$) for all $k$, then the expected growth in size between two consecutive odd integers in such a trajectory is the multiplicative factor

$$\left(\frac{3}{2}\right)^{1/2}\left(\frac{3}{4}\right)^{1/4}\left(\frac{3}{8}\right)^{1/8} \cdots = \frac{3}{4} < 1.$$

Consequently this heuristic argument suggests that on average the iterates in a trajectory tend to shrink in size, so that divergent trajectories should not exist. Furthermore it suggests that the total stopping time $\sigma_x(n)$ is (in some average sense) a constant multiple of $\log n$.

From the viewpoint of this heuristic argument, the central difficulty of the $3x + 1$ problem lies in understanding in detail the "mixing" properties of iterates of the function $T(n)$ (mod $2^k$) for all powers of 2. The function $T(n)$ does indeed have some "mixing" properties given by Theorems B and K below; these are much weaker than what one needs to settle the $3x + 1$ Conjecture.

**2.2. Behavior of the stopping time function.** It is Riho Terras's ingenious observation that although the behavior of the total stopping time function seems hard to analyze, a great deal can be said about the stopping time function. He proved the following fundamental result ([**64**], [**65**]), also found independently by Everett [**31**].

THEOREM A (Terras). *The set of integers $S_k = \{n\colon n$ has stopping time $\leqslant k\}$ has a limiting asymptotic density $F(k)$, i.e., the limit*

$$F(k) = \lim_{x \to \infty} \frac{1}{x} \#\{n\colon n \leqslant x \quad \text{and} \quad \sigma(n) \leqslant k\}$$

*exists. In addition, $F(k) \to 1$ as $k \to \infty$, so that almost all integers have a finite stopping time.*

The ideas behind Terras's analysis seem basic to a deeper understanding of the $3x + 1$ problem, so I describe them in detail. In order to do this, I introduce some notation to describe the results of the process of iterating the function $T(n)$. Given an integer $n$, define a sequence of 0-1 valued quantities $x_i(n)$ by

$$T^{(i)}(n) \equiv x_i(n) \pmod 2, \qquad 0 \leqslant i < \infty, \tag{2.2}$$

where $T^{(0)}(n) = n$. The results of the first $k$ iterations of $T$ are completely described by the *parity vector*

$$\mathbf{v}_k(n) = (x_0(n), \dots, x_{k-1}(n)), \tag{2.3}$$

since the result of $k$ iterations is

$$T^{(k)}(n) = \lambda_k(n)n + \rho_k(n), \tag{2.4}$$

where

$$\lambda_k(n) = \frac{3^{x_0(n) + \cdots + x_{k-1}(n)}}{2^k} \tag{2.5}$$

and

$$\rho_k(n) = \sum_{i=0}^{k-1} x_i(n) \frac{3^{x_{i+1}(n) + \cdots + x_{k-1}(n)}}{2^{k-i}}. \tag{2.6}$$

Note that in (2.5), (2.6) both $\lambda_k$ and $\rho_k$ are completely determined by the parity vector $\mathbf{v} = \mathbf{v}_k(n)$ given by (2.3); I sometimes indicate this by writing $\lambda_k(\mathbf{v})$, $\rho_k(\mathbf{v})$ (instead of $\lambda_k(n), \rho_k(n)$). The formula (2.4) shows that a necessary condition for $T^{(k)}(n) < n$ is that

$$\lambda_k(n) < 1, \tag{2.7}$$

since $\rho_k(n)$ is nonnegative. Terras [**64**] defines the *coefficient stopping time* $\omega(n)$ to be the least value of $k$ such that (2.7) holds, and $+\infty$ if no such value of $k$ exists. It is immediate that

$$\omega(n) \leqslant \sigma(n). \tag{2.8}$$

The function $\omega(n)$ plays an important role in the analysis of the behavior of the stopping time function $\sigma(n)$, see Theorem C.

The formula (2.2) expresses the parity vector $\mathbf{v} = \mathbf{v}_k(n)$ as a function of $n$. Terras's idea is to reverse this process and express $n$ as a function of $\mathbf{v}$.

THEOREM B. *The function* $Q_k\colon \mathbf{Z} \to \mathbf{Z}/2^k\mathbf{Z}$ *defined by*

$$Q_k(n) = \sum_{i=0}^{k-1} x_i(n)2^i$$

*is periodic with period* $2^k$. *The induced function* $\overline{Q}_k\colon \mathbf{Z}/2^k\mathbf{Z} \to \mathbf{Z}/2^k\mathbf{Z}$ *is a permutation, and its order is a power of* 2.

*Proof* (sketch). The theorem is established by induction on $k$, using the inductive hypotheses:

(1) $x_i(n)$ is periodic with period $2^{i+1}$ for $0 \leqslant i \leqslant k-1$. In fact

$$x_i(n + 2^i) \equiv x_i(n) + 1 \pmod 2 \tag{2.9}$$

for $0 \leqslant i \leqslant k-1$.

(2) $Q_k(n)$ is periodic with period $2^k$.

(3) $\lambda_k(n)$ and $\rho_k(n)$ are periodic with period $2^k$.

(4) $\overline{Q}_k$ is a permutation whose order divides $2^k$. Also

$$\overline{Q}_k(n + 2^{k-1}) \equiv \overline{Q}_k(n) + 2^{k-1} \pmod{2^k}. \tag{2.10}$$

I omit the details. ∎

The cycle structure and order of the first few permutations $\overline{Q}_k$ are given in Table 2. (One-cycles are omitted.) It is interesting to observe that the order of the permutation $\overline{Q}_k$ seems to be much smaller than the upper bound $2^k$ proved in Theorem B. Is there some explanation of this phenomenon?

TABLE 2. Cycle structure and order of permutation $\overline{Q}_k$.

| $k$ | $\overline{Q}_k$ | order |
|---|---|---|
| 1 | identity | 1 |
| 2 | identity | 1 |
| 3 | (1,5) | 2 |
| 4 | (1,5)(2,10)(9,13) | 2 |
| 5 | (1,21)(2,10)(4,20)(5,17)(7,23)(9,29,25,13)(18,26) | 4 |
| 6 | (1,21)(2,42)(3,35)(4,20)(5,17,37,49) (7,23)(8,40)(9,29,25,13)(10,34) (18,58,50,26)(19,51)(27,59)(33,53) (36,52)(39,55)(41,61,57,45) | 4 |

Theorem B allows one to associate with each vector $\mathbf{v} = (v_0, \dots, v_{k-1}) \in (\mathbf{Z}/2\mathbf{Z})^k$ of length $k$ a unique congruence class $S(\mathbf{v}) \pmod{2^k}$ given by

$$S(\mathbf{v}) = \{ n: \mathbf{v} = (x_0(n), \dots, x_{k-1}(n)) \}.$$

The integer

$$n_0(\mathbf{v}) \equiv (\overline{Q}_k)^{-1}\left( \sum_{i=0}^{k-1} v_i 2^i \right) \qquad \pmod{2^k}$$

with $0 \leqslant n_0(\mathbf{v}) < 2^k$ is the minimal element in $S(\mathbf{v})$ and $S(\mathbf{v})$ is the arithmetic progression:

$$S(\mathbf{v}) = \{ n_0(\mathbf{v}) + 2^k i: 0 \leqslant i < \infty \}.$$

Now I consider the relation between a vector $\mathbf{v}$ and stopping times for integers $n \in S(\mathbf{v})$. Define a vector $\mathbf{v} = (v_0, v_1, \dots, v_{k-1})$ of length $k$ to be *admissible* if

(1) $(v_0 + \cdots + v_{k-1})\ln 3 < k \ln 2$,

(2) $(v_0 + \cdots + v_i)\ln 3 > (i+1)\ln 2$, when $0 \leqslant i \leqslant k-2$.

Note that all admissible vectors $\mathbf{v}$ of length $k$ have

$$v_0 + \cdots + v_{k-1} = [k\theta], \tag{2.11}$$

where $\theta = \ln 2/\ln 3 = (\log_2 3)^{-1} \approx .63093$ and $[x]$ denotes the largest integer $\leqslant x$. The following result is due to Terras.

THEOREM C (Terras). (a) *The set of integers with coefficient stopping time k are exactly the set of integers in those congruence classes n* $\pmod{2^k}$ *for which there is an admissible vector* $\mathbf{v}$ *of length k with* $n = n_0(\mathbf{v})$.

(b) *Let* $n = n_0(\mathbf{v})$ *for some vector* $\mathbf{v}$ *of length k. If* $\mathbf{v}$ *is admissible, then all sufficiently large integers congruent to n* $\pmod{2^k}$ *have stopping time k. If* $\mathbf{v}$ *is not admissible, then only finitely many integers congruent to n* $\pmod{2^k}$ *have stopping time k.*

*Proof.* The assertions made in (a) about coefficient stopping times follow from the definition of admissibility, because that definition asserts that

(i) $\lambda_k(\mathbf{v}) < 1$,

(ii) $\lambda_i(\mathbf{v}) > 1$ for $1 \leqslant i \leqslant k-1$.

To prove (b), first note that if $\mathbf{v}$ is admissible of length $k$, then

$$T^{(i)}(n) \geqslant \frac{3^{v_0 + \cdots + v_{i-1}}}{2^i} n \geqslant n \quad \text{for} \quad 1 \leqslant i \leqslant k-1,$$

and so all elements of $S(\mathbf{v})$ have stopping time at least $k$. Now define $\varepsilon_k > 0$ by

$$\varepsilon_k = 1 - \frac{3^{[k\theta]}}{2^k}, \tag{2.12}$$

where $\theta = (\log_2 3)^{-1}$, and note that (2.11) implies that

$$\varepsilon_k = 1 - \lambda_k(\mathbf{v}) = 1 - \frac{3^{v_0 + \cdots + v_{k-1}}}{2^k}$$

for all *admissible* $\mathbf{v}$. Now for $n \in S(\mathbf{v})$ for an admissible $\mathbf{v}$, (2.4) may be rewritten as

$$T^{(k)}(n) = n + (\rho_k(\mathbf{v}) - \varepsilon_k n). \tag{2.13}$$

Hence when $\mathbf{v}$ is admissible, those $n$ in $S(\mathbf{v})$ with

$$n > \varepsilon_k^{-1}\rho_k(\mathbf{v}) \tag{2.14}$$

have stopping time $k$, and $\omega(n) = \sigma(n) = k$ in this case.

Now suppose $\mathbf{v}$ is not admissible. There are two cases, depending on whether or not some initial segment $(v_0, \ldots, v_i)$ of $\mathbf{v}$ is admissible. No initial segment of $\mathbf{v}$ is admissible if and only if

$$(v_0 + \cdots + v_{i-1})\log 3 > i \log 2 \quad \text{for} \quad 1 \leqslant i \leqslant k-1, \tag{2.15}$$

and when (2.15) holds say that $\mathbf{v}$ is *inflating*. If $\mathbf{v}$ is inflating, $\lambda_k(\mathbf{v}) > 1$ so that $T^{(k)}(n) \geqslant n$ for all $n$ in $S(\mathbf{v})$ by (2.4), so that no elements of $S(\mathbf{v})$ have stopping time $k$ or less. In the remaining case $\mathbf{v}$ has an initial segment $\mathbf{w} = (v_0, v_1, \ldots, v_i)$ with $i < k - 1$ which is admissible. Now $S(\mathbf{v}) \subseteq S(\mathbf{w})$ and all sufficiently large elements of $S(\mathbf{w})$ have stopping time $i + 1 < k$ by the argument just given. ■

Theorem C asserts that the set of integers $I_k$ with a given coefficient stopping time $k$ is a set of arithmetic progressions (mod $2^k$), which has the immediate consequence that $I_k$ has the asymptotic density

$$d(I_k) = \lim_{x \to \infty} \frac{1}{x} \#\{n : n \leqslant x \text{ and } n \in I_k\}$$

which is given by

$$d(I_k) = \frac{1}{2^k} \#\{\mathbf{v} : \mathbf{v} \text{ is admissible and of length } k\}.$$

Furthermore Theorem C asserts that the set

$$S_k = \{n : n \text{ has stopping time } k\}$$

differs from $I_k$ by a finite set, so that $S_k$ also has an asymptotic density which is the same as that of $I_k$. Consequently, Theorem C implies the first part of Theorem A, that the set of all integers with stopping time at most $k$ have an asymptotic density $F(k)$ given by

$$F(k) = \sum_{\substack{\mathbf{v} \text{ admissible} \\ \text{length}(\mathbf{v}) \leqslant k}} \text{weight}(\mathbf{v}), \tag{2.16}$$

where

$$\text{weight}(\mathbf{v}) = 2^{-\text{length}(\mathbf{v})}.$$

Now the formula (2.16) can be used to prove the second part of Theorem A, and in fact to prove the stronger result that $F(k)$ approaches 1 at an exponential rate as $k \to \infty$.

THEOREM D. *For all* $k \geqslant 1$,

$$1 - F(k) = \lim_{x \to \infty} \frac{1}{x} \#\{n : n \leqslant x \text{ and } \sigma(n) > k\} \leqslant 2^{-\eta k}, \tag{2.17}$$

*where*

$$\eta = 1 - H(\theta) \approx .05004\ldots. \tag{2.18}$$

*Here* $H(x) = -x\log_2 x - (1 - x)\log_2(1 - x)$ *is the entropy function and* $\theta = (\log_2 3)^{-1}$.

*Proof.* Let $C = C_1 \cup C_2$, where

$$C_1 = \{\mathbf{v} \colon \mathbf{v} \text{ is admissible and length}(\mathbf{v}) \leqslant k\}$$

and

$$C_2 = \{\mathbf{v} \colon \mathbf{v} \text{ is inflating and length}(\mathbf{v}) = k\}.$$

Then $C$ has the property that for any binary word $\mathbf{w}$ of length $k$ there is a unique $\mathbf{v} \in C$ with $\mathbf{v}$ a prefix of $\mathbf{w}$. Now for any $\mathbf{v}$ with length $(\mathbf{v}) \leqslant k$

$$\text{weight}(\mathbf{v}) = \sum \text{weight}(\mathbf{w}),$$

where the sum is over all $\mathbf{w}$ of length $k$ for which $\mathbf{v}$ is a prefix of $\mathbf{w}$. Hence

$$\sum_{\mathbf{v} \in C} \text{weight}(\mathbf{v}) = \sum_{\text{length}(\mathbf{w}) = k} \text{weight}(\mathbf{w}) = 1.$$

From (2.16) this implies that

$$\sum_{\mathbf{v} \in C_2} \text{weight}(\mathbf{v}) = 2^{-k}|C_2| = 1 - F(k),$$

where $|C_2|$ denotes the number of vectors in $C_2$. The already proved first part of Theorem A shows that

$$1 - F(k) = \lim_{x \to \infty} \frac{1}{x} \#\{n \colon n \leqslant x \text{ and } \sigma(n) > k\},$$

so that to prove (2.17) it suffices to bound $|C_2|$ from above.

Now the definition (2.15) of an inflating vector implies that

$$C_2 \subseteq \left\{\mathbf{v} \colon \sum_{i=0}^{k-1} v_i > k\theta\right\},$$

so that

$$|C_2| \leqslant \sum_{j > k\theta} \binom{k}{j}. \tag{2.19}$$

The right side of (2.19) is just the tail of the binomial distribution. It is easily checked using Stirling's formula that for any constant $\alpha > \frac{1}{2}$ and any $\varepsilon > 0$ the bound

$$\sum_{j > k\alpha} \binom{k}{j} \leqslant k \binom{k}{[k\alpha]} \leqslant 2^{(H(\alpha) + \varepsilon)k}$$

holds for all sufficiently large $k$. With more work one can obtain the more precise estimate (Ash [**8**], Lemma 4.7.2) that for any $\alpha > \frac{1}{2}$

$$\sum_{j > k\alpha} \binom{k}{j} \leqslant 2^{H(\alpha)k},$$

which used in (2.19) implies (2.17). ∎

Theorem D cannot be substantially improved; it cannot be proved that for any $\varepsilon > 0$ we have

$$|C_2| \geqslant 2^{(H(\theta) - \varepsilon)k}$$

for all sufficiently large $k$ depending on $\varepsilon$. Hence for any $\varepsilon > 0$

$$1 - F(k) \geqslant 2^{-(\eta + \varepsilon)k}$$

holds for all sufficiently large $k$ depending on $\varepsilon$.

**2.3. What is the relation between the coefficient stopping time and the stopping time?** Theorem C shows that generally they are equal: For any fixed $k$ at most a finite number of those $n$ having coefficient stopping time $\omega(n) \leqslant k$ have $\sigma(n) \neq \omega(n)$. Terras [**64**] and later Garner [**34**] conjecture that this never occurs.

COEFFICIENT STOPPING TIME CONJECTURE. *For all $n \geqslant 2$, the stopping time $\sigma(n)$ equals the coefficient stopping time $\omega(n)$.*

The Coefficient Stopping Time Conjecture has the aesthetic appeal that if it is true, then the set of positive integers with stopping time $k$ is exactly a collection of congruence classes (mod $k$), as described by part (i) of Theorem C. Furthermore, the truth of the Coefficient Stopping Time Conjecture implies that there are no nontrivial cycles. To see this, suppose that there were a nontrivial cycle of period $k$ and let $n_0$ be its smallest element, and note that $\sigma(n_0) = \infty$. Then $T^{(i)}(n_0) > n_0$ for $1 \leqslant i \leqslant k-1$ and

$$T^{(k)}(n_0) = \lambda_k(n_0)n_0 + \rho_k(n_0) = n_0. \tag{2.20}$$

Now $\rho_k(n_0) \neq 0$ since $n_0$ isn't a power of 2, so that (2.20) implies that $\lambda_k(n_0) < 1$. Hence $\omega(n_0) \leqslant k$, so that $\omega(n_0) \neq \sigma(n_0)$.

The following result shows that the Coefficient Stopping Time Conjecture is "nearly true." I will use it later to bound the number of elements not having a finite stopping time.

THEOREM E. *There is an effectively computable constant $k_0$ such that if $\mathbf{v}$ is admissible of length $k \geqslant k_0$, then all elements of $S(\mathbf{v})$ have stopping time $k$ except possibly the smallest element $n_0(\mathbf{v})$ of $S$.*

*Proof* (sketch). The results of A. Baker and N. I. Feldman on linear forms in logarithms of algebraic numbers ([**10**], Theorem 3.1) imply that there is an effectively computable absolute constant $c_0 > 0$ such that for all $k, l \geqslant 1$,

$$|k \log 2 - l \log 3| \geqslant k^{-c_0}.$$

Consequently there is an effectively computable absolute constant $c_1$ such that for $k, l \geqslant c_1$ one has

$$|2^k - 3^l| \geqslant \frac{1}{2} 2^k k^{-c_1},$$

and (2.12) then yields

$$\varepsilon_k \geqslant k^{-c_1}.$$

Since $\mathbf{v}$ is admissible, $v_0 + \cdots + v_{k-1} \leqslant \theta k$, where $\theta = (\log_2 3)^{-1}$ by (2.11). Therefore

$$\begin{aligned} \rho_k(\mathbf{v}) &= \sum_{i=0}^{k-1} v_i \frac{3^{v_{i+1} + \cdots + v_{k-1}}}{2^{k-i}} \leqslant \left( \sum_{i=0}^{[k\theta]} \frac{3^i}{2^{i+1}} \right) + k(1-\theta)\left(\frac{3}{2}\right)^{\theta k} \\ &\leqslant k 2^{(1-\theta)k}. \end{aligned}$$

But all elements of $S(\mathbf{v})$ except $n_0(\mathbf{v})$ exceed $2^k$ and

$$2^k > k^{c_1+1} 2^{(1-\theta)k} > \varepsilon_k^{-1} \rho_k(\mathbf{v})$$

for all sufficiently large $k$, so the theorem follows by (2.14). ∎

**2.4. How many elements don't have a finite stopping time?** The results proved so far can be used to obtain an upper bound for the number of elements not having a finite stopping time. Let

$$\pi^*(x) = |\{n : n \leqslant x \text{ and } \sigma(n) < \infty\}|.$$

The following result is the sharpest known result concerning the size of the "exceptional" set of $n$ with $\sigma(n) = \infty$.

THEOREM F. *There is a positive constant* $c_1$ *such that*

$$|\pi^*(x) - x| \leqslant c_1 x^{1-\eta}, \tag{2.21}$$

*where* $\eta \approx .05004\ldots$ *is the constant defined in Theorem* D.

*Proof.* Suppose $2^{k-1} \leqslant x \leqslant 2^k$. Then

$$|\{n : n \leqslant x \text{ and } \sigma(n) = \infty\}| = |\pi^*(x) - x| \leqslant S_1 + S_2,$$

where $S_1 = \#\{n \leqslant 2^k\colon\ \omega(n) \geqslant k+1\}$ and $S_2 = \#\{n \leqslant 2^k\colon\ \omega(n) \leqslant k \text{ and } \omega(n) \neq \sigma(n)\}$. Now Theorem D shows that

$$S_1 \leqslant c_2 (2^k)^{1-\eta} \leqslant 2c_2 x^{1-\eta}, \tag{2.22}$$

and Theorem E shows that

$$S_2 \leqslant \#\{\mathbf{v}\colon \mathbf{v} \text{ admissible and length}(\mathbf{v}) \leqslant k\} + c_3,$$

where $c_3 = \#\{n\colon\ \omega(n) \leqslant k_0 \text{ and } \omega(n) \neq \sigma(n)\}$ is a constant by Theorem C. Now

$$\#\{\mathbf{v}\colon \mathbf{v} \text{ admissible and length}(\mathbf{v}) = i\} \leqslant \#\{\mathbf{v}\colon v_0 + \cdots + v_{i-1} = [i\theta]\} = \binom{i}{[i\theta]}$$
$$\leqslant c_4 2^{(1-\eta)i}$$

using the binomial theorem and Stirling's formula. Hence

$$S_2 \leqslant c_5 2^{(1-\eta)k} + c_3 \leqslant (2c_5 + c_3) x^{1-\eta}.$$

Then this inequality and (2.22) imply (2.21) with $c_1 = 2c_2 + c_3 + 2c_5$. ■

**2.5. Behavior of the total stopping time function.** Much less is known about the total stopping time function than about the stopping time function. One phenomenon immediately observable from a table of the total stopping times of small integers is the occurrence of many pairs and triples of integers having the same finite total stopping time. From Figure 1 we see that $\sigma_\infty(20) = \sigma_\infty(21) = 6$, $\sigma_\infty(12) = \sigma_\infty(13) = 7$, $\sigma_\infty(84) = \sigma_\infty(85) = 8$, $\sigma_\infty(52) = \sigma_\infty(53) = 9$, and $\sigma_\infty(340) = \sigma_\infty(341) = 10$. Indeed for larger values of $n$, multiple consecutive values occur with the same total stopping time. For example there are 17 consecutive values of $n$ with $\sigma_\infty(n) = 40$ for $7083 \leqslant n \leqslant 7099$. A related phenomenon is that over short ranges of $n$ the function $\sigma_\infty(n)$ tends to assume only a few values (C. W. Dodge [**67**]). As an example the values of $\sigma_\infty(n)$ for $1000 \leqslant n \leqslant 1099$ are given in Table 3. Only 19 values for $\sigma_\infty(n)$ are observed, for which a frequency count is given in Table 4. Both of these phenomena have a simple explanation; they are caused by coalescence of trajectories of different $n$'s after a few steps. For example the trajectories of $8k+4$ and $8k+5$ coalesce after 3 steps, for all $k \geqslant 0$. More generally, the large number of coalescences of numbers $n_1$ and $n_2$ close together in size can be traced to the trivial cycle (1,2), as follows. Suppose $n_1$ and $n_2$ have $\sigma_\infty(n_1) \equiv \sigma_\infty(n_2) \pmod 2$, and let $\sigma_\infty(n_1) = r_1 \geqslant \sigma_\infty(n_2) = r_2$. Then the trajectories of $n_1$ and $n_2$ coalesce after at most $r_1 - 1$ iterations, since $T^{(r_1-1)}(n_1) = T^{(r_1-1)}(n_2) = 2$, since the trajectory of $n_2$ continues to cycle around the trivial cycle. If in addition $\lambda_{r_1-1}(n_1) = \lambda_{r_1-1}(n_2)$, which nearly always happens if $n_1$ and $n_2$ are about the same size, then the trajectories of $2^{r_1-1}k + n_1$, and $2^{r_1-1}k + n_2$ coalesce after at most $r_1 - 1$ iterations, for $k \geqslant 0$. In particular, $\sigma_\infty(2^{r_1-1}k + n_1) = \sigma_\infty(2^{r_1-1}k + n_2)$ then holds for $k \geqslant 1$. In this case the original coalescence of $n_1$ and $n_2$ has produced an infinite arithmetic progression $(\operatorname{mod} 2^{r_1-1})$ of coalescences. The gradual accumulation of all these arithmetic progressions of coalescences of numbers close together in size leads to the phenomena observed in Tables 3 and 4.

Although the $3x+1$ Conjecture asserts that all integers $n$ have a finite total stopping time, the strongest result proved so far concerning the density of the set of integers with a finite total stopping time is much weaker.

TABLE 3. Values of the total stopping time $\sigma_\infty(n)$ for $1000 \leqslant n \leqslant 1099$.

| | 1000 -1009 | 1010 -1019 | 1020 -1029 | 1030 -1039 | 1040 -1049 | 1050 -1059 | 1060 -1069 | 1070 -1079 | 1080 -1089 | 1090 -1099 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 72 | 42 | 34 | 80 | 23 | 23 | 80 | 18 | 31 | 31 |
| 1 | 91 | 42 | 34 | 26 | 80 | 61 | 80 | 107 | 88 | 31 |
| 2 | 72 | 72 | 42 | 80 | 80 | 53 | 80 | 23 | 31 | 23 |
| 3 | 29 | 72 | 42 | 99 | 80 | 53 | 50 | 18 | 88 | 23 |
| 4 | 45 | 26 | 10 | 80 | 23 | 53 | 23 | 18 | 31 | 23 |
| 5 | 45 | 26 | 26 | 80 | 23 | 107 | 50 | 18 | 31 | 50 |
| 6 | 45 | 34 | 26 | 80 | 80 | 23 | 23 | 23 | 88 | 61 |
| 7 | 61 | 99 | 26 | 80 | 80 | 53 | 42 | 23 | 88 | 88 |
| 8 | 72 | 34 | 80 | 42 | 23 | 23 | 18 | 34 | 15 | 61 |
| 9 | 72 | 42 | 80 | 42 | 42 | 23 | 18 | 34 | 31 | 23 |

TABLE 4. Values of $\sigma_\infty(n)$ and their frequencies for $1000 \leqslant n \leqslant 1099$.

| $\sigma_\infty(n)$ | freq. | $\sigma_\infty(n)$ | freq. | $\sigma_\infty(n)$ | freq. | $\sigma_\infty(n)$ | freq. |
|---|---|---|---|---|---|---|---|
| 10 | 1 | 29 | 1 | 50 | 3 | 88 | 5 |
| 15 | 1 | 31 | 7 | 53 | 4 | 91 | 1 |
| 18 | 6 | 34 | 6 | 61 | 4 | 99 | 2 |
| 23 | 17 | 42 | 9 | 72 | 6 | 107 | 2 |
| 26 | 6 | 45 | 3 | 80 | 16 | | |

THEOREM G (Crandall). *Let*

$$\pi_{\text{total}}(x) = |\{n : n \leqslant x \text{ and } \sigma_\infty(n) < \infty\}|.$$

*Then there is a positive constant* $c_4$ *such that*

$$\pi_{\text{total}}(x) > x^{c_4}$$

*for all sufficiently large* $x$.

Assuming that the $3x + 1$ Conjecture is true, one can consider the problem of determining the *expected size* of the total stopping time function $\sigma_\infty(n)$. Crandall **[28]** and Shanks **[60]** were guided by probabilistic heuristic arguments (like the one described earlier) to conjecture that the average order of $\sigma_\infty(n)$ should be a constant times $\ln n$; more precisely, that

$$\frac{1}{x}\sum_{n=1}^{x} \sigma_\infty(n) \sim 2\left(\ln\frac{4}{3}\right)^{-1} \ln x.$$

A modest amount of empirical evidence supports these conjectures, see **[28]**.

**2.6. Are there non-trivial cycles?** A first observation is that there are other cycles if negative integers are allowed in the domain of the function. There is a cycle of period 1 starting from $n = -1$, and there are cycles of length 3 and 11 starting from $n = -5$ and $n = -17$, respectively. Böhm and Sontacchi **[13]** conjecture that these cycles together with the cycles starting with $n = 0$ and $n = 1$ make up the entire set of cycles occurring under iteration of $T(n)$ applied to the integers $\mathbf{Z}$. Several authors have proposed the following conjecture (**[13]**, **[28]**, **[41]**, **[64]**).

FINITE CYCLES CONJECTURE. *There are only a finite number of distinct cycles for the function* $T(n)$ *iterated on the domain* $\mathbf{Z}$.

One can easily show that for any given length $k$ there are only a finite number of integers $n$ that are periodic under iteration by $T$ with period $k$, in fact at most $2^k$ such integers, as observed by Böhm and Sontacchi **[13]**. To see this, substitute the equation (2.4) into

$$T^{(k)}(n) = n, \quad n \in \mathbf{Z} \tag{2.23}$$

to obtain the equation

$$\left(1 - \frac{3^{x_0 + \cdots + x_{k-1}}}{2^k}\right) n = \frac{3^{x_0 + \cdots + x_{k-1}}}{2^k} \sum_{i=0}^{k-1} x_i \frac{2^i}{3^{x_0 + \cdots + x_i}}. \tag{2.24}$$

There are only $2^k$ choices for the 0-1 vector $\mathbf{v} = (x_0, \ldots, x_{k-1})$, and for each choice of $\mathbf{v}$ the equation (2.24) determines a unique rational solution $n = n(\mathbf{v})$. Consequently there are at most $2^k$ solutions to (2.23). Böhm and Sontacchi also noted that this gives an (inefficient) finite procedure for deciding if there are any cycles of a given length $k$, as follows: Determine the rational number $n(\mathbf{v})$ for each of the $2^k$ vectors $\mathbf{v}$, and for each $n(\mathbf{v})$ which is an integer test if (2.23) holds.

The argument of Böhm and Sontacchi is a very general one that makes use only of the fact that the necessary condition (2.24) for a cycle has a unique solution when the values $x_i$ are fixed. In fact, considerably more can be proved about the nonexistence of nontrivial cyclic trajectories using special features of the necessary condition (2.24). For example, several authors have independently found a much more efficient computational procedure for proving the nonexistence of nontrivial cyclic trajectories of period $\leqslant k$; it essentially makes use of the inequality

$$(1 - \lambda_k(\mathbf{v}))\, n \leqslant \rho_k(\mathbf{v}),$$

which must hold for $\mathbf{v} = (x_0, x_1, \ldots, x_k)$ satisfying (2.24). This approach also allows one to check the truth of the Coefficient Stopping Time Conjecture for all $n$ with $\omega(n) \leqslant k$. The basic result is as follows.

THEOREM H (Terras). *For each $k$ there is a finite bound $M(k)$ given by*

$$M(k) = \max\{\varepsilon_i^{-1}\rho_i(\mathbf{v}) : \mathbf{v} \text{ admissible, length } (\mathbf{v}) = i \leqslant k\} \tag{2.25}$$

*such that $\omega(n) \leqslant k$ implies that $\omega(n) = \sigma(n)$ whenever $n \geqslant M(k)$. Consequently*:

(i) *If $\sigma(n) < \infty$ for all $n \leqslant M(k)$, then there are no non-trivial cycles of length $\leqslant k$.*

(ii) *If $\omega(n) = \sigma(n)$ for all $n \leqslant M(k)$, then $\omega(n) \leqslant k$ implies $\omega(n) = \sigma(n)$.*

*Proof*. The existence of the bound $M(k)$ follows immediately from (2.14), and (ii) follows immediately from this fact.

To prove (i), suppose a nontrivial cycle of length $\leqslant k$ exists. We observed earlier that if $n_0$ is the smallest element in a purely periodic nontrivial cycle of length $\leqslant k$, then $\omega(n_0) = i \leqslant k$ and $\sigma(n_0) = \infty$. The first part of the theorem then implies that $n_0 \leqslant M(k)$. This contradicts the hypothesis of (i). ■

Theorem H can be used to show the nonexistence of nontrivial cycles of small period by obtaining upper bounds for the $M(k)$ and checking that condition (i) holds. This approach has been taken by Crandall [**28**], Garner [**34**], Schuppar [**58**] and Terras [**64**]. In estimating $M(k)$, one can show that the quantities $\rho_i(\mathbf{v})$ are never very large, so that the size of $M(k)$ is essentially determined by how large

$$\varepsilon_i^{-1} = \left(1 - \frac{3^{[i\theta]}}{2^i}\right)^{-1}$$

can get. The worst cases occur when $3^{[i\theta]}$ is a very close approximation to $2^i$, i.e., when $i/[i\theta]$ is a very good rational approximation to $\phi = \log_2 3$. The best rational approximations to $\phi$ are given by the convergents $p_k/q_k$ of the continued fraction expansion of $\phi = [1; 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3 \ldots]$. Crandall [**28**] uses general properties of continued fraction convergents to obtain the following quantitative result.

THEOREM I (Crandall). *Let $n_0$ be the minimal element of a purely periodic trajectory of period $k$. Then*

$$k > \frac{3}{2}\min\left(q_j, \frac{2n_0}{q_j + q_{j+1}}\right), \tag{2.26}$$

*where $p_i/q_j$ is any convergent of the continued fraction expansion of* $\log_2 3$ *with* $j \geqslant 4$.

As an application, use Yoneda's bound [**2**] that $n_0 > 2^{40}$ and choose $j = 13$ in (2.26), noting that $q_{13} = 190737$ and $q_{14} = 10590737$, to conclude that *there are no nontrivial cycles with period length less than* 275,000.

Further information about the nonexistence of nontrivial cyclic trajectories can be obtained by treating the necessary condition (2.24) as an *exponential Diophantine equation*. Davidson [**29**] calls a purely periodic trajectory of period $k$ a *circuit* if there is a value $i$ for which

$$n_0 < T(n_0) < \cdots < T^{(i)}(n_0)$$

and

$$T^{(i)}(n_0) > T^{(i+1)}(n_0) > \cdots > T^{(k)}(n_0) = n_0,$$

i.e., the parity vector $\mathbf{v}_k(n_0) = (x_0(n_0), \dots, x_{k-1}(n_0))$ has the special form

$$x_j(n_0) = \begin{cases} 1, & \text{when } 0 \leqslant j \leqslant [k\theta] - 1, \\ 0, & \text{when } [k\theta] \leqslant j \leqslant k - 1, \end{cases} \tag{2.27}$$

where $\theta = (\log_2 3)^{-1}$. The cycle starting with $n_0 = 1$ is a circuit. Davidson observes that each solution to the exponential Diophantine equation

$$(2^{a+b} - 3^b)h = 2^a - 1, a \geqslant 1 \tag{2.28}$$

gives rise to a circuit of length $k = a + b$ with $[k\theta] = b$ and $n_0 = 2^b h - 1$, and conversely. (The equation (2.28) is the necessary condition (2.24) specialized to the vector (2.27).) R. Steiner [**61**] showed that $(a, b, h) = (1, 1, 1)$ is the only solution of (2.28), thus proving the following result.

THEOREM J (Steiner). *The only cycle that is a circuit is the trivial cycle.*

*Proof* (sketch). Steiner's method is to show first that any solution of (2.28) with $a \geqslant 4$ has the property that $(a + b)/b$ is a convergent in the continued fraction expansion of $\log_2 3$, since (2.28) implies that

$$0 < \left| \frac{a+b}{b} - \log_2 3 \right| \leqslant \frac{1}{b \ln 2(2^b - 1)}. \tag{2.29}$$

He checks that this rational approximation $(a + b)/b$ is so good that it violates the effective estimates of A. Baker [**10**, p. 45] for linear forms in logarithms of algebraic numbers if $b > 10^{199}$. Finally he checks that (2.29) fails to hold for all that $b < 10^{199}$ by computing the convergents of the continued fractions of $\log_2 3$ up to $10^{199}$. ∎

The most remarkable thing about Theorem J is the weakness of its conclusion compared to the strength of the methods used in its proof. The proof of Theorem J does have the merit that it shows that the coefficient Stopping Time Conjecture holds for the *infinite set* of admissible vectors $\mathbf{v}$ of the form (2.27).

**2.7. Do divergent trajectories exist?** Several authors have observed that heuristic probabilistic arguments suggest that no divergent trajectories occur.

DIVERGENT TRAJECTORIES CONJECTURE. *The function $T$: $\mathbf{Z} \to \mathbf{Z}$ has no divergent trajectories, i.e., there exists no integer $n_0$ for which*

$$\lim_{k \to \infty} |T^{(k)}(n_0)| = \infty. \tag{2.30}$$

If a divergent trajectory $\{T^{(k)}(n_0)\colon 0 \leqslant k < \infty\}$ exists, it cannot be equidistributed (mod 2). Indeed if one defines

$$N^*(k) = |\{ j\colon j \leqslant k \text{ and } T^{(j)}(n_0) \equiv 1 \pmod 2\}|,$$

then it can be proved that the condition (2.30) implies that

$$\liminf_{k\to\infty} \frac{N^*(k)}{k} \geqslant (\log_2 3)^{-1} \approx .63097. \tag{2.31}$$

Theorem F constrains the possible behavior of divergent trajectories. Indeed, associated to any divergent trajectory $D = \{T^{(k)}(n_0)\colon k \geqslant 1\}$ is the infinite set $U_D = \{n\colon n \in D \text{ and } T^{(k)}(n) > n \text{ for all } k \geqslant 1\}$. Since $\sigma(n) = \infty$ for all $n \in U_D$, Theorem F implies that

$$|\{n \in U_D\colon n \leqslant x\}| \leqslant c_1 x^{1-\eta}, \tag{2.32}$$

where $\eta \approx .05004$. Roughly speaking, (2.32) asserts that the elements of a divergent trajectory cannot go to infinity "too slowly."

**2.8. Connections of the $3x + 1$ problem to ergodic theory.** The study of the general behavior of the iterates of measure preserving functions on a measure space is called *ergodic theory*. The $3x + 1$ problem has some interesting connections to ergodic theory, because the function $T$ extends to a measure-preserving function on the 2-adic integers $\mathbf{Z}_2$ defined with respect to the 2-adic measure. To explain this, I need some basic facts about the 2-adic integers $\mathbf{Z}_2$, cf. **[14]**, **[49]**. The 2-adic integers $\mathbf{Z}_2$ consist of all series

$$\alpha = a_0 + a_1 2 + a_2 2^2 + \cdots, \qquad \text{all } a_i = 0 \text{ or } 1,$$

where the $\{a_i\colon 0 \leqslant i < \infty\}$ are called the *2-adic digits* of $\alpha$. One can define congruences $(\operatorname{mod} 2^k)$ on $\mathbf{Z}_2$ by $\alpha \equiv \beta \pmod{2^k}$ if the first $k$ 2-adic digits of $\alpha$ and $\beta$ agree. Addition and multiplication on $\mathbf{Z}_2$ are given by

$$X = \alpha + \beta \Leftrightarrow X \,(\operatorname{mod} 2^k) \equiv \alpha(\operatorname{mod} 2^k) + \beta(\operatorname{mod} 2^k) \text{ for all } k,$$

$$X = \alpha\beta \Leftrightarrow X\,(\operatorname{mod} 2^k) \equiv \alpha(\operatorname{mod} 2^k) \cdot \beta(\operatorname{mod} 2^k) \text{ for all } k.$$

The *2-adic valuation* $|\ |_2$ on $\mathbf{Z}_2$ is given by $|0|_2 = 0$ and for $\alpha \neq 0$ by $|\alpha|_2 = 2^{-k}$, where $a_k$ is the first nonzero 2-adic digit of $\alpha$. The valuation $|\ |_2$ induces a metric $d$ on $\mathbf{Z}_2$ defined by

$$d(\alpha, \beta) = |\alpha - \beta|_2.$$

As a topological space $\mathbf{Z}_2$ is compact and complete with respect to the metric $d$; a basis of open sets for this topology is given by the *2-adic discs of radius* $2^{-k}$ about $\alpha$:

$$B_k(\alpha) = \{\beta \in \mathbf{Z}_2\colon \alpha \equiv \beta \pmod{2^k}\}.$$

Finally one may consistently define the *2-adic measure* $\mu_2$ on $\mathbf{Z}_2$ so that

$$\mu_2(B_k(\alpha)) = 2^{-k};$$

in particular $\mu_2(\mathbf{Z}_2) = 1$. The integers $\mathbf{Z}$ are a subset of $\mathbf{Z}_2$; for example

$$-1 = 1 + 1 \cdot 2 + 1 \cdot 2^2 + \cdots.$$

Now one can extend the definition of the function $T\colon \mathbf{Z} \to \mathbf{Z}$ given by (2.1) to $T\colon \mathbf{Z}_2 \to \mathbf{Z}_2$ by

$$T(\alpha) = \begin{cases} \dfrac{\alpha}{2}, & \text{if } \alpha \equiv 0 \pmod 2, \\[2mm] \dfrac{3\alpha + 1}{2}, & \text{if } \alpha \equiv 1 \pmod 2. \end{cases}$$

Ergodic theory is concerned with the extent to which iterates of a function mix subsets of a measure space. I will use the following basic concepts of ergodic theory specialized to the measure space $\mathbf{Z}_2$ with the measure $\mu_2$. A measure-preserving function $H\colon \mathbf{Z}_2 \to \mathbf{Z}_2$ is *ergodic* if the only $\mu_2$-measurable sets $E$ for which $H^{-1}(E) = E$ are $\mathbf{Z}_2$ and the empty set, i.e., such a function does such a good job of mixing points in the space that it has no nontrivial $\mu_2$-invariant sets. It can be shown **[39**, p. 36] that an equivalent condition for ergodicity is that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \mu_2(H^{-j}(B_k(\alpha)) \cap B_l(\beta)) = \mu_2(B_k(\alpha))\mu_2(B_l(\beta)) = 2^{-(k+l)},$$

for all $\alpha, \beta \in \mathbf{Z}_2$ and all integers $k, l \geqslant 0$. This condition in turn is equivalent to the assertion that for almost all $\alpha \in \mathbf{Z}_2$ the sequence of iterates

$$\{H^i(\alpha): i = 0, 1, 2, \dots\}$$

is uniformly distributed $(\bmod 2^k)$ for all $k \geqslant 1$. A function $H: \mathbf{Z}_2 \to \mathbf{Z}_2$ is *strongly mixing* if

$$\lim_{N \to \infty} \mu_2\big(H^{-N}(B_k(\alpha)) \cap B_l(\beta)\big) = 2^{-(k+l)}$$

for all $\alpha, \beta \in \mathbf{Z}_2$ and all $k, l \geqslant 0$. Strongly mixing functions are ergodic.

The following result is a special case of a result of K. P. Matthews and A. M. Watts [**50**].

THEOREM K. *The map $T$ is a measure-preserving transformation of $\mathbf{Z}_2$ which is strongly mixing. Consequently it is ergodic, and hence for almost all $\alpha \in \mathbf{Z}_2$ the sequence*

$$\{T^{(i)}(\alpha): i = 0, 1, 2, \dots\}$$

*is uniformly distributed* $(\bmod 2^k)$ *for all $k \geqslant 1$.*

Theorem K implies nothing about the behavior of $T$ on the set of integers $\mathbf{Z}$ because it is a measure 0 subset of $\mathbf{Z}_2$. In fact, the trajectory $\{T^{(i)}(n): i = 0, 1, 2, \dots\}$ of any integer $n$ can *never* have the property of the conclusion of Theorem K, for if the trajectory is eventually periodic with period $k$, it cannot be uniformly distributed $(\bmod 2^{k+1})$, while if it is a divergent trajectory, it cannot even be equidistributed $(\bmod 2)$ by (2.31). Consequently, this connection of the $3x + 1$ problem to ergodic theory does not seem to yield any deep insight into the $3x + 1$ problem itself.

There is, however, another connection of the $3x + 1$ problem to ergodic theory on $\mathbf{Z}_2$ that may conceivably yield more information on the $3x + 1$ problem. For each $\alpha \in \mathbf{Z}_2$ define the 0-1 variables $x_i$ by

$$T^{(i)}(\alpha) \equiv x_i \ (\bmod 2).$$

Now define the function $Q_\infty: \mathbf{Z}_2 \to \mathbf{Z}_2$ by $Q_\infty(\alpha) = \beta$, where

$$\beta = x_0 + x_1 2 + x_2 2^2 + \cdots . \tag{2.33}$$

The value $Q_\infty(\alpha)$ thus encodes the behavior of *all* the iterates of $\alpha$ under $T$. The following result has been observed by several people, including R. Terras and C. Pomerance, but has not been explicitly stated before.

THEOREM L. *The map $Q_\infty: \mathbf{Z}_2 \to \mathbf{Z}_2$ is a continuous, one-one, onto, and measure-preserving map on the 2-adic integers $\mathbf{Z}_2$.*

*Proof.* This is essentially a consequence of Theorem B. Use the fact that $Q_\infty(\alpha) \equiv \overline{Q}_n(\alpha)$ $(\bmod 2^n)$. For any $\alpha_1, \alpha_2$ in $\mathbf{Z}_2$, if $|\alpha_1 - \alpha_2|_2 \leqslant 2^{-n}$, then $\alpha_1 \equiv \alpha_2 \ (\bmod 2^n)$, so

$$Q_\infty(\alpha_1) \equiv \overline{Q}_n(\alpha_1) \equiv \overline{Q}_n(\alpha_2) \equiv Q_\infty(\alpha_2) \ (\bmod 2^n),$$

so that $|Q_\infty(\alpha_1) - Q_\infty(\alpha_2)| \leqslant 2^{-n}$ and $Q_\infty$ is continuous. If $\alpha_1 \neq \alpha_2$, then $\alpha_1 \not\equiv \alpha_2 \ (\bmod 2^n)$ for some $n$, so that

$$Q_\infty(\alpha_1) \equiv \overline{Q}_n(\alpha_1) \not\equiv \overline{Q}_n(\alpha_2) \equiv Q_\infty(\alpha_2) \ (\bmod 2^n)$$

and $Q_\infty$ is one-to-one. To see that $Q_\infty$ is onto, given $\alpha$ one can find $\beta_n$ so that

$$\overline{Q}_n(\beta_n) \equiv \alpha \ (\bmod 2^n),$$

since $\overline{Q}_n$ is a permutation. Then $|Q_\infty(\beta_n) - \alpha|_2 \leqslant 2^{-n}$. Now $\{\beta_n\}$ forms a Cauchy sequence in the 2-adic metric and $\mathbf{Z}_2$ is compact, hence the limiting value $\beta$ of $\{\beta_n\}$ satisfies $Q_\infty(\beta) = \alpha$. Now $Q_\infty^{-1}$ is defined, and $Q_\infty(\alpha) \equiv \overline{Q}_n^{-1}(\alpha) \ (\bmod 2^n)$ implies that $Q_\infty^{-1}$ is continuous. ∎

The $3x + 1$ Conjecture can be reformulated in terms of the function $Q_\infty$ as follows.

$3x + 1$ CONJECTURE (Third form). *Let $\mathbf{N}^+$ denote the positive integers. Then $Q_\infty(\mathbf{N}^+) \subseteq \frac{1}{3}\mathbf{Z}$. In fact $Q_\infty(N^+) \subseteq \frac{1}{3}\mathbf{Z} - \mathbf{Z}$.*

For example $Q_\infty(1) = \Sigma_{i=0}^{\infty} 2^{2n} = -1/3, Q_\infty(2) = -2/3$, and $Q_\infty(3) = -20/3$.

The behavior of the function $Q_\infty$ under iteration is itself of interest. Let $\mathbf{Q}_2$ denote the set of all rational numbers having odd denominators, so that $\mathbf{Q}_2 \subseteq \mathbf{Z}_2$. The set $\mathbf{Q}_2$ consists of exactly those 2-adic integers whose 2-adic expansion is finite or eventually periodic. The Finite Cycles Conjecture is equivalent to the assertion that there is a finite odd integer $M$ such that

$$Q_\infty(\mathbf{Z}) \subseteq \frac{1}{M}\mathbf{Z}.$$

In fact one can take $M = \Pi(2^l - 1)$, where the product runs over all integers $l$ for which there is a cycle of minimal length $l$. As a hypothesis for further work I advance the following conjecture.

PERIODICITY CONJECTURE. $Q_\infty(\mathbf{Q}_2) = \mathbf{Q}_2$.

For example, one may calculate that $Q_\infty(10) = -26/3$, $Q_\infty(-26/3) = -54$, $Q_\infty(-54) = -82/7$, $Q_\infty(-82/7) = ?/15$. It can be shown that if $n$ has a divergent trajectory, then the sequence $(x_0(n), x_1(n), x_2(n), \dots)$ cannot be eventually periodic. As a consequence the truth of the Periodicity Conjecture implies the truth of the Divergent Trajectories Conjecture.

Theorem B has a curious consequence concerning the fixed points of iterates of $Q_\infty$.

THEOREM M. *Suppose the kth iterate $Q_\infty^{(k)}$ of $Q_\infty$ has a fixed point $\alpha \in \mathbf{Z}_2$ which is not a fixed point of any $Q_\infty^{(l)}$ for $1 \leqslant l < k$. Then $k$ is a power of 2.*

*Proof.* By hypothesis $Q_\infty^{(k)}(\alpha) = \alpha$ and $Q_\infty^{(l)}(\alpha) = \alpha_l \neq \alpha$, for $1 \leqslant l < k$. All the $\alpha_l$'s are distinct for $0 \leqslant l \leqslant k$, since $Q_\infty^{(l_1)}(\alpha) = Q_\infty^{(l_2)}(\alpha)$ implies $Q_\infty^{(l_1 - l_2)}(\alpha) = \alpha$, since $Q_\infty$ is one-one and onto. Consequently one can pick $m$ large enough so that all the residue classes $\alpha_l \pmod{2^m}$ are distinct, for $0 \leqslant l \leqslant k$, where $\alpha_0 = \alpha$. Now the action of $Q_\infty \pmod{2^m}$ is exactly that of the permutation $\overline{Q}_m$, hence

$$\overline{Q}_m^{(l)}(\alpha \pmod{2^m}) \equiv \alpha_l \pmod{2^m}$$

for $0 \leqslant l < k$. In particular $(\alpha_0 \pmod{2^m}, \alpha_1 \pmod{2^m}, \dots, \alpha_{k-1} \pmod{2^m})$ makes up a single cycle of the permutation $\overline{Q}_m$, hence $k$ is a power of 2 by Theorem B. ∎

**3. Generalizations of the $3x + 1$ problem.** The $3x + 1$ problem can be generalized by considering other functions $U$: $\mathbf{N} \to \mathbf{N}$ defined on the natural numbers $\mathbf{N}$ that are similar to the function $T$. The functions I consider to be similar to the function $T$ are the *periodically linear functions*, which are those functions $U$ for which there is a finite modulus $d$ such that the function $U$ when restricted to any congruence class $k \pmod d$ is linear. Some reasons to study generalizations of the $3x + 1$ problem are that they may uncover new phenomena, they can indicate the limits of validity of known results, and they can lead to simpler, more revealing proofs. Here I discuss three directions of generalizations of the $3x + 1$ problem. These deal with algorithmic decidability questions, with the existence of stopping times for almost all integers, and with the fractional parts of $(3/2)^k$.

**3.1. Algorithmic decidability questions.** J. H. Conway [26] proved the remarkable result that a simple generalization of the $3x + 1$ problem is algorithmically undecidable. He considers the class $\mathbf{F}$ of periodically piecewise linear functions $g$: $\mathbf{N} \to \mathbf{N}$ having the structure

$$g(n) = \frac{a_k}{(k, d)} n \text{ if } n \equiv k \pmod d, \quad \text{for } 0 \leqslant k \leqslant d - 1, \tag{3.1}$$

specified by the nonnegative integers $(d, \alpha_0, \dots, \alpha_{d-1})$. These are exactly the functions $g$: $\mathbf{N} \to \mathbf{N}$ such that $g(n)/n$ is periodic.

THEOREM O (Conway). *For every partial recursive function $f$ defined on a subset $D$ of the natural numbers $\mathbf{N}$ there exists a function $g$: $\mathbf{N} \to \mathbf{N}$ such that*

(1) *$g(n)/n$ is periodic $\pmod d$ for some $d$ and takes rational values.*

(2) *There is some iterate* $k \geqslant 1$ *such that* $g^{(k)}(n) = 2^j$ *for some* $j$ *if and only if* $n$ *is in* $D$.
(3) $g^{(k)}(n) = 2^{f(n)}$ *for the minimal* $k \geqslant 1$ *such that* $g^{(k)}(n)$ *is a power of* 2.

Conway's proof actually gives in principle a procedure for explicitly constructing such a function $g$ given a description of a Turing machine* that computes $f$. He carried out this procedure to find a function $g$ associated to a particular partial recursive function $f$ having the property that $f(2^{p_n}) = p_{n+1}$, where $p_n$ is the $n$th prime; this is described in Guy [**38**].

By choosing a particular partial recursive function whose domain is not a recursive subset of **N**, e.g., a function $f_0$ that encodes the halting problem for Turing machines, we obtain the following corollary of Theorem O.

THEOREM P (Conway). *There exists a particular, explicitly constructible function* $g_0\colon \mathbf{N} \to \mathbf{N}$ *such that* $g_0(n)/n$ *is periodic* (mod $d$) *for a finite modulus* $d$ *and takes rational values, for which there is no Turing machine that, when given* $n$, *always decides in a finite number of steps whether or not some iterate* $g_0^{(k)}(n)$ *with* $k \geqslant 1$ *is a power of* 2.

**3.2. Existence of stopping times for almost all integers.** Several authors have investigated the range of validity of the result that $T(n)$ has a finite stopping time for almost all integers $n$ by considering more general classes of periodically linear functions. One such class **G** consists of all functions $U = U(m, d, R)$ which are given by

$$U(n) = \begin{cases} \dfrac{n}{d}, & \text{if } n \equiv 0 \pmod{d}, \\[2mm] \dfrac{mn - r}{d}, & \text{if } n \not\equiv 0 \pmod{d}, \text{ and } r \in R \text{ is such that} \\ & mn \equiv r \pmod{d}, \end{cases} \tag{3.2}$$

where $m$ and $d$ are positive integers with $(m, d) = 1$ and $R = \{r_i\colon r_i \equiv i \pmod{d}, 1 \leqslant i \leqslant d-1\}$ is a fixed set of residue class representatives of the nonzero residue classes (mod $d$). The $3x+1$ function $T$ is in the class **G**. H. Möller [**52**] completely characterized the functions $U = U(m, d, R)$ in the set **G** which have a finite stopping time for almost all integers $n$. He showed they are exactly those functions for which

$$m < d^{d/(d-1)}. \tag{3.3}$$

E. Heppner [**41**] proved the following quantitative version of this result, thereby generalizing Theorem D.

THEOREM Q (Heppner). *Let* $U = U(m, d, R)$ *be a function in the class* **G**.
(i) *If* $m < d^{d/(d-1)}$, *then there exist real numbers* $\delta_1, \delta_2 > 0$ *such that for* $N = [\log x/\log d]$ *we have* $\#\{n\colon n \leqslant x \text{ and } U^{(N)}(n) > nx^{-\delta_1}\} = O(x^{1-\delta_2})$ *as* $x \to \infty$.
(ii) *If* $m > d^{d/(d-1)}$, *then there exist real numbers* $\delta_3, \delta_4 > 0$ *such that for* $N = [\log x/\log d]$ *we have* $\#\{n\colon n \leqslant x \text{ and } U^{(N)}(n) < nx^{\delta_3}\} = O(x^{1-\delta_4})$ *as* $x \to \infty$.

J.-P. Allouche [**1**] has further sharpened Theorem Q and Matthews and Watts [**53**], [**54**] have extended it to a larger class of functions.

It is a measure of the difficulty of problems in this area that even the following apparently weak conjecture is unsolved.

EXISTENCE CONJECTURE. *Let* $U$ *be any function in the class* **G**. *Then*:
(i) $U$ *has at least one purely periodic trajectory if* $m < d^{d/(d-1)}$;
(ii) $U$ *has at least one divergent trajectory if* $m > d^{d/(d-1)}$.

**3.3. Fractional parts of $(3/2)^k$.** Attempts to understand the distribution (mod 1) of the sequence $\{(3/2)^k\colon 1 \leqslant k < \infty\}$ have uncovered oblique connections with ergodic-theoretic aspects of a generalization of the $3x+1$ problem. It is conjectured that the sequence $(3/2)^k$ is uniformly

*Conway's proof used Minsky machines, which have the same computational power as Turing machines.

distributed (mod 1). (This conjecture seems intractable at present.) One approach to this problem is to determine what kinds of (mod 1) distributions can occur for sequences $\{(3/2)^k\xi\colon 1 \leqslant k < \infty\}$, where $\xi$ is a fixed real number. In this vein K. Mahler [**48**] considered the problem of whether or not there exist real numbers $\xi$, which he called *Z-numbers*, having the property that

$$0 \leqslant \left\{\left(\frac{3}{2}\right)^k \xi\right\} \leqslant \frac{1}{2}, k = 1,2,3,\ldots, \tag{3.4}$$

where $\{x\} = x - [x]$ is the fractional part of $x$. He showed that the set of $Z$-numbers is countable, by showing that there is at most one $Z$-number in each interval $[n, n+1)$, for $n = 1,2,3,\ldots$. He went on to show that a necessary condition for the existence of a $Z$-number in the interval $[n, n+1)$ is that the trajectory $(n, W(n), W^{(2)}(n), \ldots)$ of $n$ produced by the periodically linear function

$$W(n) = \begin{cases} \dfrac{3n}{2}, & \text{if } n \equiv 0 \pmod 2, \\ \dfrac{3n+1}{2}, & \text{if } n \equiv 1 \pmod 2, \end{cases} \tag{3.5}$$

satisfy

$$W^{(k)}(n) \not\equiv 3 \pmod 4, \qquad 1 \leqslant k < \infty. \tag{3.6}$$

Mahler concluded from this that it is unlikely that any $Z$-numbers exist. This is supported by the following heuristic argument. The function $W$ may be interpreted as acting on the 2-adic integers by (3.5), and it has properties exactly analogous to the properties of $T$ given by Theorem K. In particular, for almost all 2-adic integers $\alpha$ the sequence of iterates $(\alpha, W(\alpha), W^{(2)}(\alpha), \ldots)$ has infinitely many values $k$ with $W^{(k)}(\alpha) \equiv 3 \pmod 4$. Thus if a given $n \in \mathbf{Z}$ behaves like almost all 2-adic integers $\alpha$, then (3.6) will not hold for $n$. Note that it is possible that all the trajectories $(n, W(n), W^{(2)}(n), \ldots)$ for $n \geqslant 1$ are uniformly distributed (mod $2^k$) for *all* $k$, unlike the behavior of the function $T(n)$.

In passing, I note that the possible distributions (mod 1) of $\{(3/2)^k\xi;\ 1 \leqslant k < \infty\}$ for real $\xi$ have an intricate structure (see G. Choquet [**16–22**] and A. D. Pollington [**55, 56**]). In particular, Pollington [**56**] proves that there are uncountably many real numbers $\xi$ such that

$$\frac{1}{25} \leqslant \left\{\left(\frac{3}{2}\right)^k \xi\right\} \leqslant \frac{24}{25};\ k = 1,2,3,\ldots,$$

in contrast to the at most countable number of solutions $\xi$ of (3.4).

**4. Conclusion.** *Is the* $3x + 1$ *problem intractably hard*? The difficulty of settling the $3x + 1$ problem seems connected to the fact that it is a deterministic process that simulates "random" behavior. We face this dilemma: On the one hand, to the extent that the problem has structure, we can analyze it—yet it is precisely this structure that seems to prevent us from proving that it behaves "randomly". On the other hand, to the extent that the problem is structureless and "random", we have nothing to analyze and consequently cannot rigorously prove anything. Of course there remains the possibility that someone will find some hidden regularity in the $3x + 1$ problem that allows some of the conjectures about it to be settled. The existing general methods in number theory and ergodic theory do not seem to touch the $3x + 1$ problem; in this sense it seems intractable at present. Indeed all the conjectures made in this paper seem currently to be out of reach if they are true; I think there is more chance of disproving those that are false.

*If the* $3x + 1$ *problem is intractable*, *why should one bother to study it*? One answer is provided by the following aphorism: "No problem is so intractable that something interesting cannot be said about it." Study of the $3x + 1$ problem has uncovered a number of interesting phenomena; I believe further study of it may be rewarded by the discovery of other new phenomena. It also serves as a benchmark to measure the progress of general mathematical theories. For example,

future developments in solving exponential Diophantine equations may lead to the resolution of the Finite Cycles Conjecture.

*If all the conjectures made in this paper are intractable, where would one begin to do research on this deceptively simple problem*? As a guide to doing research, I ask questions. Here are a few that occur to me: For the $3x + 1$ problem, what restrictions are there on the growth in size of members of a divergent trajectory assuming that one exists? What interesting properties does the function $Q_\infty$ have? Is there some direct characterization of $Q_\infty$ other than the recursive definition (2.33)?

## References


An asterisk (*) denotes a research paper on the $3x + 1$ problem and its generalizations.

***1.** J.-P. Allouche, Sur la conjecture de "Syracuse-Kakutani-Collatz," Séminaire de Théorie des Nombres, 1978–1979, Exp. No. 9, 15, pp., CNRS, Talence (France) 1979. MR 81g:10014.

**2.** Shiro Ando, letter to J. C. Lagarias, Feb. 18, 1983. Reports that Prof. Nabuo Yoneda (Dept. of Information Science, Tokyo Univ.) has verified the $3x + 1$ conjecture for all $n < 2^{40} = 1.2 \times 10^{12}$.

**3.** R. V. Andree, Modern Abstract Algebra, Holt, Reinhart, and Winston, 1971.

**4.** Anon., The $3x + 1$ problem, Popular Computing, 1 (April 1973) 1–2.

**5.** Anon., $3x + 1$ (continued), Popular Computing, 1 (July 1973) 6–7.

**6.** Anon., $3x + 1$ strings, Popular Computing, 2 (April 1973) 12–13.

**7.** Anon., $3x + 1$ once again, Popular Computing, 3 (April 1975) 4–5.

**8.** R. Ash, Information Theory, Interscience Publishers, Wiley, New York, 1965.

**9.** A. O. L. Atkin, Comment on Problem 63-13, SIAM Review, 8 (1966) 234–236.

**10.** A. Baker, Transcendental Number Theory, Cambridge Univ. Press, 1975.

**11.** M. Beeler, W. Gosper, R. Schroeppel, HAKMEM, Memo 239, Artificial Intelligence Laboratory, M.I.T. 1972, p. 64. Presents some numerical results on cycles for the $3x + 1$ and related problems.

***12.** J. Blazewicz and A. Pettorossi, Collatz's Conjecture and Binary Sequences, Institute of Control Engineering, Technical University of Poznan, Poznan, Poland (preprint) March 1983.

***13.** C. Böhm and G. Sontacchi, On the existence of cycles of given length in integer sequences like $x_{n+1} = x_n/2$ if $x_n$ even, and $x_{n+1} = 3x + 1$ otherwise, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., 64 (1978) 260–264, MR 83h.10030.

**14.** Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.

**15.** M. J. Bruce, Crazy roller coaster, Math. Teacher, 71 (January 1978) 45–49.

**16.** G. Choquet, Répartition des nombres $k\ (3/2)^n$; mesures et ensembles associés, C. R. Acad. Sci. Paris, 290 (31 mars 1980) 575–580

**17.** ______, Algorithmes adaptés aux suites $(k\theta^n)$ et aux chaînes associées, C. R. Acad. Sci. Paris, 290 (28 avril 1980) 719–724.

**18.** ______, $\theta$-Jeux récursifs et application aux suites $(k\theta^n)$; solénoïdes de $\Pi^Z$, C. R. Acad. Sci. Paris, 290 (19 mai 1980) 863–868.

**19.** ______, Construction effective de suites $(k(3/2)^n)$. Etude des mesures (3/2)-stables, C. R. Acad. Sci. Paris, 291 (15 septembre 1980) 69–74.

**20.** ______, Les fermés (3/2)-stables de $\Pi$; structure des fermés dénombrables; applications arithmétiques, C. R. Acad. Sci. Paris, 291 (29 septembre 1980) 239–244.

**21.** ______, $\theta$-fermés; $\theta$-chaînes et $\theta$-cycles (pour $\theta = 3/2$), C. R. Acad. Sci. Paris, 292 (5 janvier 1981) 5–10.

**22.** ______, $\theta$-fermés et dimension de Hausdorff. Conjectures de travail. Arithmétique des $\theta$-cycles (où $\theta = 3/2$), C. R. Acad. Sci. Paris, 292 (9 février 1981) 339–344. See especially Conjecture D, which arises from a generalization of the problem considered by Mahler [**48**]; cf. (3.4) in the text.

**23.** L. Collatz, letters to R. Terras, 7 May 1976 and 2 Sept. 1976.

**24.** ______, letter to L. Garner, 17 March 1980. In none of these letters does L. Collatz actually state that he proposed the $3x + 1$ problem.

**25.** ______, Verzweigungsdiagramme und Hypergraphen, International Series for Numerical Mathematics, vol. 38, Birkhäuser, 1977. This paper remarks on the use of graphical representations to study iteration problems, as in Fig. 1 in the text. It does not discuss the $3x + 1$ problem.

***26.** J. H. Conway, Unpredictable Iterations, Proc. 1972 Number Theory Conference, University of Colorado, Boulder, Colorado (1972) 49–52.

**27.** H. S. M. Coxeter, Cyclic Sequences and Frieze Patterns (The Fourth Felix Behrend Memorial Lecture), 1970. In this lecture Coxeter offered a $50 prize for a proof of the $3x + 1$ Conjecture and $100 for a counterexample, according to C. W. Trigg [**67**].

***28.** R. E. Crandall, On the "$3x + 1$" problem, Math. Comp., 32 (1978) 1281–1292.

*29. J. L. Davidson, Some Comments on an Iteration Problem, Proc. 6th Manitoba Conf. on Numerical Mathematics (1976) 155–159.

30. R. Dunn, On Ulam's Problem, Dept. of Computer Science Report #CU-CS-011-73, University of Colorado, Boulder, January 1973. Reports on computer programs to test the $3x + 1$ Conjecture and to calculate the densities $F(k)$ in (2.16) for $k \leqslant 21$.

*31. C. J. Everett, Iteration of the number theoretic function $f(2n) = n$, $f(2n + 1) = 3n + 2$, Adv. Math., 25 (1977) 42–45.

32. A. S. Fraenkel, private communication 1981.

33. M. Gardner, Mathematical Games, Scientific American, 226 (June 1972) 114–118.

*34. L. E. Garner, On the Collatz $3n + 1$ algorithm, Proc. Amer. Math. Soc., 82 (1981) 19–22.

35. F. Gruenberger, $3x + 1$ revisited, Popular Computing, 7 (Oct. 1979) 3–12.

36. R. K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981 (Problem E16.)

37. ______, Don't try to solve these Problems!, this MONTHLY, 90 (1983) 35–41.

*38. ______, Conway's prime producing machine, Math. Mag., 56 (1983) 26–33.

39. P. Halmos, Lectures on Ergodic Theory, Math. Soc. of Japan, Tokyo, 1956 (Reprint: Chelsea, New York).

40. H. Hasse, Unsolved Problems in Elementary Number Theory, Lectures at U. Maine (Orono), Spring 1975, dittoed notes, pp. 23–33. He states (p. 23) that Thompson has proved the Finite Cycles Conjecture, a result I am unable to verify.

*41. E. Heppner, Eine Bemerkung zum Hasse-Syracuse Algorithmus, Archiv. Math., 31 (1978) 317–320, MR 80d. 10007.

42. I. N. Herstein and I. Kaplansky, Matters Mathematical, 2nd ed., Chelsea, New York, 1978, 44–45. They call the problem the "Syracuse algorithm."

43. D. Hofstadter, Gödel, Escher, Bach, Basic Books, New York, 1979. Brings the $3x + 1$ problem to a wide audience by stating it on pp. 400–402.

44. S. D. Isard and A. M. Zwicky, Three open questions in the theory of one-symbol Smullyan systems, SIGACT News (1969) 11–14. They propose the problem: Let $f(n) = n/3$ if $n \equiv 0 \pmod 3$ and let $g(n) = 2n + 1$. Can all numbers $n \equiv 0$ or $1 \pmod 3$ be reduced to 1 by a series of operations $f$ and $g$? See [67], p. 148.

45. S. Kakutani, private communication 1981.

46. David Kay, Undergraduate Research Project no. 2, Pi Mu Epsilon J. (1972) 338.

47. M. S. Klamkin, Problem 63-12, SIAM Review, 5 (1963) 275–276. States the problem: Consider the permutation $f\colon N^+ \to N^+$ defined by $f(3n) = 2n$, $f(3n - 1) = 4n - 1$, $f(3n - 2) = 4n - 3$. Does $n = 8$ have an infinite trajectory? How many finite cycles are there for $f$? This function was the original one proposed by L. Collatz in 1932 [24].

*47a. G. M. Leigh and K. R. Matthews, A generalization of the Syracuse algorithm to $F_q[x]$, preprint.

*48. K. Mahler, An unsolved problem on the powers of 3/2, J. Austral. Math. Soc., 8 (1968) 313–321. The problem Mahler studies was originally proposed by Prof. Saburô Uchiyama (Tsukuba Univ.), according to S. Ando [2]. It arose indirectly in connection with the function $g(k)$ of Waring's problem, see G. H. Harding and E. M. Wright, An Introduction to the Theory of Numbers (4th ed.), Oxford Univ. Press, 1960, Theorem 393ff.

49. ______, $P$-adic numbers and their functions, 2nd ed., Cambridge Univ. Press, 1976.

*50. K. R. Matthews and A. M. Watts, A generalization of Hasse's generalization of the Syracuse algorithm, Acta Arith., 43 (1983) 75–83.

*51. ______, A Markov approach to the generalized Syracuse algorithm, preprint.

*51a. ______, Some invariant probability measures on the polyadic numbers, preprint.

*52. H. Möller, Über Hasses Verallgemeinerung des Syracuse-Algorithmus (Kakutani's Problem), Acta Arith., 34 (1978) 219–226. This paper claims A. Fraenkel checked that $f^{(k)}(n) = 1$ for all $n \leqslant 2^{50}$, a result in error [32]. In a note added in proof it asserts the proofs of Terras [64] are faulty. They seem essentially correct to me, and Terras [65] supplies elided details.

53. J. Nievergelt, J. Farrar, and E. M. Reingold, Computer Approaches to Mathematical Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974, 211–217. They discuss the problem of checking the $3x + 1$ Conjecture for a given $n$ and give some interesting empirical observations.

54. C. Stanley Ogilvy, Tomorrow's Math, Oxford Univ. Press, Oxford, 1972, 103.

55. A. D. Pollington, Intervals Constructions in the Theory of Numbers, Thesis, University of London, 1980. See especially Chapter 5, which contains results related to Mahler's problem [48].

56. ______, Progressions arithmétiques géneralisées et le problème des $(3/2)^n$, C. R. Acad. Sci. Paris, 292 (16 février 1981) 383–384.

57. L. Ratzan, Some work on an unsolved palindromic algorithm, Pi Mu Epsilon J., 5 (Fall 1973) 463–466.

*57a. D. A. Rawsthorne, Imitation of an Iteration, preprint.

*58. B. Schuppar, Kettenbrüche und der $3a + 1$ Algorithmus, preprint.

**59.** D. Shanks, Comments on Problem 63–13, SIAM Review, 7 (1965) 284–286. Gives the result of a computer search on Collatz' original problem (see [9]), [47]), and a heuristic concerning lengths of finite cycles.

**60.** ______, Western Number Theory Conference Problem 1975 #5 (R. Guy, ed.) (mimeographed sheets). States a conjecture on the average size of the total stopping time function.

***61.** R. P. Steiner, A Theorem on the Syracuse Problem, Proc. 7th Manitoba Conference on Numerical Mathematics–1977, Winnipeg (1978) 553–559.

***62.** ______, On the "$Qx + 1$ Problem," $Q$ odd, Fibonacci Quart., 19 (1981) 285–288.

***63.** ______, On the "$Qx + 1$ Problem," $Q$ odd II, Fibonacci Quart., 19 (1981) 293–296.

***64.** R. Terras, A stopping time problem on the positive integers, Acta Arith. 30 (1976) 241–252.

***65.** ______, On the existence of a density, Acta Arith., 35 (1979) 101–102.

**66.** R. Tijdeman, Det Kongelige Norske Videnskabers Selskak, Skrifter, 16, 1972.

**67.** C. W. Trigg, C. W. Dodge, and L. F. Meyers, Comments on Problem 133, Eureka (now Crux Mathematicorum) 2, no. 7 (August–September 1976) 144–150. Problem 133 is the $3x + 1$ problem, proposed by K. S. Williams, who said he was shown it by one of his students. C. W. Trigg gives a good history of the problem.

**68.** M. Vaughan-Lee, letter to L. Garner, Aug. 10, 1981. Michael Vaughan-Lee (Oxford U.) has verified the $3x + 1$ Conjecture for $n < 2^{32} \cong 4.5 \times 10^9$ on a Sinclair 2X81 Microcomputer.

**69.** C. Williams, B. Thwaites, A. van der Poorten, W. Edwards, and L. Williams, Ulam's Conjecture continued again, 1000 pounds for proof, Rumored proof disclaimed, References, PPC Calculator J., 9 (Sept. 1982) 23–24. Gives some history of the problem. B. Thwaites states he originated the $3x + 1$ problem in 1952 and offers 1000 pounds for a proof of the $3x + 1$ Conjecture.

**70.** M. Yamada, A convergence proof about an integral sequence, Fibonacci Quart., 18 (1980) 231–242. The proofs in this paper are incorrect, see MR 82d:10026.

# PERPENDICULAR POLYGONS

J. CHRIS FISHER, D. RUOFF, AND J. SHILLETO

*Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, Canada S4S 0A2*

**1. Introduction.** We present a theory to deal with geometric problems of a varied nature. The apparently disparate concepts of Fourier analysis and Hermitian forms are fundamental to the theory. Indeed, our exposition might serve as an introduction to these important notions since it relies only on the elementary properties of complex numbers, vectors, and polynomials.

Our approach evolved from the decomposition theory of $n$-gons initiated by Bachmann and Schmidt [1], which in turn has its roots in Fourier analysis: an $n$-gon is written as a linear combination of "regular" $n$-gons, with the scalars being the associated finite Fourier coefficients. Typical examples to illustrate the theory are provided in Section 2. Many of these have been discussed already in an earlier paper [12]. The deeper metric properties of polygons then are approached through Hermitian forms on $n$-tuples of complex numbers. In Section 4 Hermitian forms are applied to isoperimetric inequalities for polygonal regions. This leads us in the final section to consider regions bounded by continuous curves. For this we extend our results concerning finite dimensional vector spaces to the corresponding results in Hilbert space. The finite Fourier series extend quite naturally to their more familiar infinite cousins.

---

All three authors came to Regina during the late 60's when the former college was expanding into a major university. Chris Fisher received his Ph.D. from the University of Toronto (1972) under the supervision of H. S. M. Coxeter. He has also been visiting Professor at the University of Bologna, the Technical University of Munich, and the Free University of Brussels. On the basis of his contacts, he has enriched substantially the guest speaker program of his comparatively remote university. Dieter Ruoff studied at the University of Zurich. He wrote his Ph.D. thesis under the guidance of B. L. van der Waerden on a topic in the foundations of geometry, and he has worked in this field since then. Before emigrating to Canada he spent two postdoctoral years with F. Bachmann in Kiel. He now specializes in incidence geometry. Like Fisher, he often visits Europe pursuing professional and personal contacts and his interests in the arts. J. Shilleto, a Berkeley Ph.D., has chosen geometry as his primary interest after beginning his career in mathematical logic and doing some topology. He believes strongly in a developmental approach to teaching mathematics and is currently engaged in writing a geometry text which reflects this view.

## 2. Affine $n$-gon Theory

*The Vector Space of n-gons.* An $n$-gon $\mathscr{A}$ in the Euclidean plane is an ordered $n$-tuple of points, $\mathscr{A} = (a_0, a_1, \dots, a_{n-1})$. These points are called *vertices*, and the line segments joining $a_0$ to $a_1$, $a_1$ to $a_2, \dots, a_{n-1}$ to $a_0$ are called *sides*. Note that we do not require an $n$-gon to be convex. We also allow any number of the vertices to coincide; in fact the $n$-fold repeated origin $O = (0, 0, \dots, 0)$ turns out to be an important $n$-gon.

By virtue of our identification of the vertex

$$a_j = (x_j, y_j)$$

with the complex number

$$a_j = x_j + iy_j \quad \left(x_j, y_j \in \mathbb{R},\ i = \sqrt{-1}\right),$$

an $n$-gon in the plane is at the same time a vector in $\mathbb{C}^n$ (that is, an $n$-tuple of complex numbers). Thus a *linear combination of n-gons* $\mathscr{A} = (a_j)_{j<n}$ and $\mathscr{B} = (b_j)_{j<n}$ is identified with the linear combination of the corresponding vectors in $\mathbb{C}^n$:

$$a\mathscr{A} + b\mathscr{B} = \left(aa_j + bb_j\right)_{j<n}, \quad a, b \in \mathbb{C}.$$

In other words, $a\mathscr{A} + b\mathscr{B}$ is the $n$-gon whose vertices are obtained as linear combinations of the vertices of $\mathscr{A}$ and $\mathscr{B}$. So $(re^{i\theta})\mathscr{A}$ is the image of $\mathscr{A}$ under a dilatation by the real factor $r$ and a rotation through the angle $\theta$, both about the origin. Also, for example, $(1/2)\mathscr{A} + (1/2)\mathscr{B}$ is the $n$-gon whose $j$th vertex is the midpoint of the segment joining the $j$th vertices of $\mathscr{A}$ and $\mathscr{B}$.

*Multiplication of n-gons by Polynomials.* Note that the identification of $n$-gons with vectors compels us to define two $n$-gons to be <u>equal</u> just in case they are equal as vectors in $\mathbb{C}^n$. For example, $(a_0, a_1, \dots, a_{n-1})$ and $(a_1, a_2, \dots, a_0)$ are distinct $n$-gons (unless $a_0 = a_1 = \cdots = a_{n-1}$) even though they have the same set of vertices and edges.

*Notation.* Let $x^k\mathscr{A}$ be the $n$-gon obtained from $\mathscr{A}$ by a cyclic shift of the entries of $(a_0, a_1, \dots, a_{n-1})$ $k$ places to the left. That is,

$$\begin{aligned} x\mathscr{A} &= (a_1, a_2, \dots, a_{n-1}, a_0), \\ x^2\mathscr{A} &= (a_2, a_3, \dots, a_0, a_1), \\ &\text{etc.} \end{aligned}$$

In matrix form we have that

$$x\mathscr{A} = (a_0, a_1, \dots, a_{n-1})\chi, \quad \text{where } \chi = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Extending the multiplication above to arbitrary polynomials $p(x) = c_0 + c_1x + \cdots + c_kx^k$ ($c_j \in \mathbb{C}$) we define $p(x)\mathscr{A}$ to be the linear combination $c_0\mathscr{A} + c_1x\mathscr{A} + \cdots + c_kx^k\mathscr{A}$.

It is also possible to describe $p(x)\mathscr{A}$ as $(a_0, a_1, \dots, a_{n-1})\Pi$ where $\Pi = c_0\chi^0 + c_1\chi + \cdots + c_k\chi^k$ is an $n \times n$ *circulant matrix* (i.e., a square matrix $(c_{ij})$ for which $c_{ij} = c_{(i+1)(j+1)}$; see [**8**] or [**9**]). However, the polynomial notation is more familiar and readily brings out its inherent geometrical features. We shall see this demonstrated in the examples to follow.

LEMMAS AND THEOREMS.

(2.1) *An n-gon* $\mathscr{A} = (a_j)_{j<n}$ *whose vertices repeat after m steps satisfies* $(x^m - 1)\mathscr{A} = O$ (*where, as always*, $O = (0, 0, \dots, 0)$).

Indeed, $(x^m - 1)\mathscr{A} = O$ means precisely $a_{j+m} - a_j = 0$ for all $j = 0, 1, \ldots, n-1$. As an important special case, $(x^n - 1)\mathscr{A} = O$ *for every n-gon* $\mathscr{A}$. It follows that $x^{n-k}\mathscr{A}$ represents a shift of vertices $k$ places to the right. We usually write this as $x^{-k}\mathscr{A}$. In the same spirit,

$$\frac{1}{n}(x^{n-1} + x^{n-2} + \cdots + 1)\mathscr{A} = \frac{1}{n}\left(\frac{x^n - 1}{x - 1}\right)\mathscr{A}$$

is the $n$-gon whose vertices all coincide with the centroid of $\mathscr{A}$.

We define the $n$-gon $\mathscr{A} = (a_0, a_1, \ldots, a_{n-1})$ to be *k-regular* if there exists a point such that the rotation through the angle $2k\pi/n$ about it moves $a_0$ to $a_1$, $a_1$ to $a_2, \ldots, a_{n-1}$ to $a_0$ (see Fig. 2A).

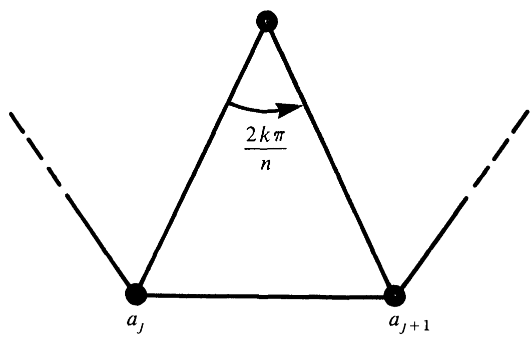



FIG. 2A. A $k$-regular $n$-gon.

Throughout this paper we denote the $n$th roots of unity as powers of

$$w = e^{2\pi i/n}.$$

Then, in our shift notation, we have the following result.

(2.2) $\mathscr{A}$ *is a k-regular n-gon whose centroid is* 0 *if and only if there is an integer k for which*

$$(x - w^k)\mathscr{A} = O.$$

Although one usually calls such $n$-gons "regular" [**6**, p. 36], we employ the prefix $k$ to emphasize that each side subtends an angle of $2k\pi/n$ at the center. When $k = 1$, $\mathscr{A}$ is convex and labelled counterclockwise; when $k = -1$, $\mathscr{A}$ is convex and clockwise. For $k$ relatively prime to $n$, $\mathscr{A}$ is a regular star $n$-gon; generally speaking it is some regular $(n/d)$-gon ($d =$ the greatest common divisor of $n$ and $k$) described $d$ times. We shall denote by $\mathscr{R}_k$ the special $k$-regular $n$-gon centered at 0 with $a_0 = 1$, namely,

$$\mathscr{R}_k = (1, w^k, w^{2k}, \ldots, w^{(n-1)k}).$$

The image of a $k$-regular $n$-gon under an affinity is called *k-affinely regular*. Recall that an affinity is the product of a nonsingular linear transformation and a translation [**6**, §13.3]. An affinity therefore preserves ratios of lengths of parallel line segments. Affinely regular polygons are discussed in [**5**], [**6**, §13.4], [**12**], and [**23**]. It is a simple geometric fact (see Fig. 2B) that a $k$-regular $n$-gon $\mathscr{A}$ centered at 0 fulfills

$$\frac{1}{2}(a_j + a_{j+2}) = \cos(2k\pi/n)\, a_{j+1}. \tag{2.3}$$

This condition leads to the following characterization.

(2.4) *An n-gon* $\mathscr{A}$ *is k-affinely regular with centroid* 0 *if and only if for some integer k*

$$(x^2 - c_k x + 1)\mathscr{A} = O,$$

*where* $c_k = 2\cos(2k\pi/n)$.

Note that (2.4) allows the vertices of $\mathscr{A}$ to be collinear; this is somewhat nonstandard but is convenient for our algebraic approach. We are obliged, however, to prove that *when the vertices of* $\mathscr{A}$ *are not collinear*, (2.4) *characterizes the n-gons with centroid at* 0 *that are images of k-regular n-gons under affinities.*

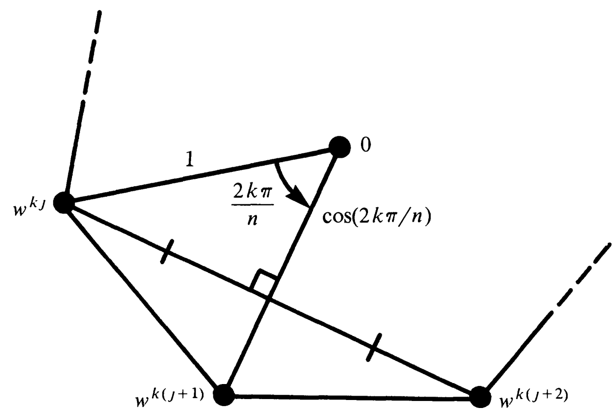



FIG. 2B. Justifying (2.4): $\frac{1}{2}(w^{kj} + w^{k(j+2)}) = \cos(2k\pi/n)w^{k(j+1)}$.

*Proof.* Since (2.3) is preserved by affinities that fix the origin, it (and its equivalent formulation $(x^2 - c_k x + 1)\mathscr{A} = O$) is a necessary condition. Conversely, one defines the affinity $\alpha$ so that it maps the centroid and first two vertices of $\mathscr{R}_k$ to the corresponding points of $\mathscr{A}$. Then one uses (2.3) inductively to prove that the remaining vertices of $\mathscr{R}_k$ and $\mathscr{A}$ correspond under $\alpha$. □

Here are two familiar special cases. *All triangles are affinely regular* since a triangle with centroid 0 satisfies $(x^2 + x + 1)\mathscr{A} = O - c_1 = -1$. Also, the *affinely regular quadrangles are parallelograms* because $(x^2 + 1)\mathscr{A} = O$ holds if and only if $\mathscr{A}$ is a parallelogram with centroid $O - c_1 = 0$.

To see how our notation works, let us prove the well-known fact that *the midpoints of the sides of a quadrangle form a parallelogram:* To simplify matters let $\mathscr{A}$ be a quadrangle whose centroid is 0; that is, $\frac{1}{4}(x^3 + x^2 + x + 1)\mathscr{A} = O$. Its midpoint figure $\mathscr{B}$ satisfies $\mathscr{B} = \frac{1}{2}(x + 1)\mathscr{A}$. Thus $O = \frac{1}{2}(x^3 + x^2 + x + 1)\mathscr{A} = (x^2 + 1)[\frac{1}{2}(x + 1)\mathscr{A}] = (x^2 + 1)\mathscr{B}$, which shows that $\mathscr{B}$ is a parallelogram (with centroid 0).

The next theorem was inspired by [**10**, p. 63, #57], which dealt with the case of a triangle.

(2.5) THEOREM. *Let squares be erected externally on the sides of an affinely regular n-gon* $\mathscr{A}$ *whose centroid is* 0. *Then the segment joining neighboring vertices of adjacent squares is perpendicular to the corresponding radius vector of* $\mathscr{A}$ *and is* $2 - c_k$ *times as long.* (The quantity $c_k$ is the parameter $2\cos(2k\pi/n)$ of (2.4).)

*Proof.* Labelling as in Fig. 2C, we have that

$$b_j - a_j = -i(a_{j+1} - a_j) \quad \text{and} \quad c_j - a_j = -i(a_j - a_{j-1}).$$

So

$$\mathscr{B} - \mathscr{A} = -i(x - 1)\mathscr{A},$$
$$\mathscr{C} - \mathscr{A} = -i(1 - x^{-1})\mathscr{A} \qquad (\text{by (2.1) } x^{-1}\mathscr{A} = x^{n-1}\mathscr{A}),$$

and therefore

$$\mathscr{B} - \mathscr{C} = -i(x - 2 + x^{-1})\mathscr{A} = -i\big((x - c_k + x^{-1}) - (2 - c_k)\big)\mathscr{A} = i(2 - c_k)\mathscr{A}.$$

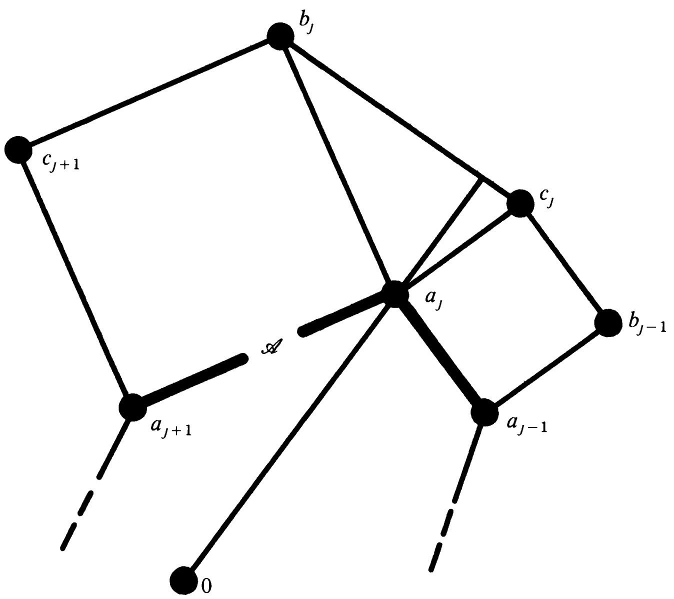



FIG. 2C. Theorem (2.5): 0 is the centroid of $\mathscr{A}$; squares are erected externally on its sides, and neighboring vertices of adjacent squares are joined.

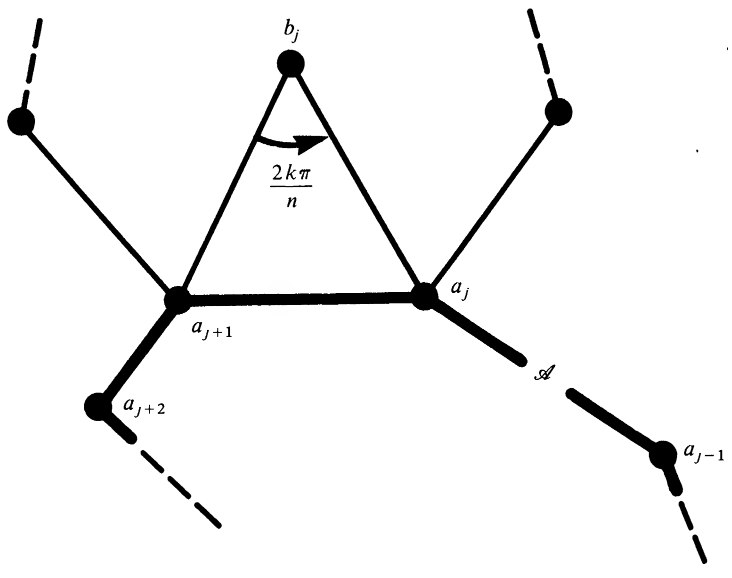



FIG. 2D. The Napoleon-Barlotti theorem: $b_j$ is the center of a $k$-regular $n$-gon erected externally on a side of $\mathscr{A}$.

The last equality uses $(x - c_k + x^{-1})\mathscr{A} = O$ which is essentially (2.4). Thus $\mathscr{B} - \mathscr{C} = i(2 - c_k)\mathscr{A}$, which is precisely what was to be proved. □

(2.6) THEOREM (Napoleon-Barlotti). *For a given n-gon $\mathscr{A}$, let $\mathscr{B}$ be the n-gon whose vertices are the centroids of k-regular n-gons erected externally on the sides of $\mathscr{A}$. Then $\mathscr{B}$ is k-regular if and only if $\mathscr{A}$ is k-affinely regular.*

*Proof.* We may assume $\mathscr{A}$ has centroid 0. By virtue of (2.2), (2.4), and the fact that $(x - w^k)(x - w^{-k}) = x^2 - 2\cos(2k\pi/n)x + 1$, we need only to show that $(x - w^k) \times (x - w^{-k})\mathscr{A} = O$ if and only if $(x - w^k)\mathscr{B} = O$. As shown in Fig. 2D, the construction of the

vertices $b_j$ of $\mathscr{B}$ may be restated by the equation

$$w^k(a_{j+1} - b_j) = a_j - b_j \qquad (j = 0, \dots, n-1,\ w = e^{2\pi i/n}).$$

More simply,

$$w^k(x\mathscr{A} - \mathscr{B}) = \mathscr{A} - \mathscr{B}.$$

So

$$\text{(a)} \qquad (x - w^{-k})\mathscr{A} = (1 - w^{-k})\mathscr{B},$$

and hence $(x - w^k)(x - w^{-k})\mathscr{A} = O$ if and only if $(x - w^k)\mathscr{B} = O$. □

This result with $n = 3$, when $\mathscr{A}$ and $\mathscr{B}$ are triangles, is popularly attributed to Napoleon, whose role in the story is a source of speculation. An extensive search (by V. G. Cavallaro [**4**]) uncovered a 1917 Italian textbook in which the triangle problem was introduced with the phrase "Theorem proposed for demonstration by Napoleon to Lagrange" (Exercise 494, page 186 of A. Faifofer, *Elementi di Geometria*, 20th ed., Sorteni e Vidotti, Venezia, 1917; it can be found also in some earlier editions such as the 18th (1912)). In addition to this reference the search turned up the result, unaccompanied by any mention of Napoleon, in an 1843 text. It was not until 1955 that the theorem for $n > 3$ (and $k = 1$) was formulated (by Barlotti in [**2**]). There have been several recent related results as well as an extension of the theorem to more general settings (see [**12**], [**13**], [**23**]).

The $n$-gon $\mathscr{C}$ of centroids of $k$-regular $n$-gons erected *internally* on the sides of the $k$-affinely regular $n$-gon $\mathscr{A}$ satisfies the equation

$$\text{(b)} \qquad (x - w^k)\mathscr{A} = (1 - w^k)\mathscr{C}.$$

Equivalently, $\mathscr{C}$ can be considered the $n$-gon of centroids of $(n-k)$-regular $n$-gons erected externally on the sides of $\mathscr{A}$. We shall call $\mathscr{B}$ and $\mathscr{C}$ the *Barlotti n-gons of* $\mathscr{A}$. Subtracting (b) from (a) we obtain that

$$\mathscr{A} = \frac{1 - w^{-k}}{w^k - w^{-k}}\mathscr{B} - \frac{1 - w^k}{w^k - w^{-k}}\mathscr{C},$$

and so

$$\text{(2.7)} \qquad \mathscr{A} = \frac{1}{w^k + 1}\mathscr{B} + \frac{w^k}{w^k + 1}\mathscr{C} \qquad (w^k \neq -1).$$

Note that when $\mathscr{A}$ is $k$-affinely regular, (2.6) implies that $(w^k + 1)^{-1}\mathscr{B}$ and $w^k(w^k + 1)^{-1}\mathscr{C}$ are $k$- and $(n-k)$-regular respectively; thus we have proved that *every $k$-affinely regular $n$-gon is the sum of a $k$-regular $n$-gon and an $(n-k)$-regular $n$-gon*. This fact is only the tip of the iceberg.

The fundamental result in [**23**], which is a clarification and simplification of the main theorem of Bachmann and Schmidt [**1**, §6.2], says that an $n$-gon can be written as a linear combination of "simpler" $n$-gons. For us the following special case of this theorem will be important.

(2.8) $\mathscr{R}_0, \mathscr{R}_1, \dots, \mathscr{R}_{n-1}$ *form a basis for the vector space* $\mathbb{C}^n$ *of n-gons over the complex numbers.*

This is actually a restatement of a well-known fact about finite Fourier series (see [**18**] or [**25**], for example). More precisely, for an $n$-gon $\mathscr{A} = (a_j)_{j<n}$ there exists a unique set of $n$ complex numbers $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$, the *finite Fourier coefficients of* $\mathscr{A}$, such that $\mathscr{A} = \sum_{k<n} \alpha_k \mathscr{R}_k$. In terms of the vertices of $\mathscr{A}$ this means that

$$\text{(2.8}'\text{)} \qquad a_j = \sum_{k<n} \alpha_k w^{jk} \qquad (w = e^{2\pi i/n}).$$

It is a standard argument ([**18**, §1] or [**25**, §1]) to prove (2.8) by showing that each $\alpha_k$ satisfies

$$\text{(2.9)} \qquad \alpha_k = \frac{1}{n}\sum_{j<n} a_j w^{-jk}.$$

Note that, in particular, $\alpha_0$ is the centroid of $\mathscr{A}$.

**3. Hermitian Forms.** We shall prove that signed area is the quadratic form associated with some Hermitian form on the vector space $\mathbb{C}^n$ of all $n$-gons. This will permit proofs about areas using the techniques developed above.

Let us denote by $I(\mathscr{A},\mathscr{B})$ the standard inner product $\Sigma_{j<n} a_j \bar{b}_j$ defined on $\mathbb{C}^n$. ($I$ is used to suggest inertia since $I(\mathscr{A},\mathscr{A})$ is the moment of inertia of $\mathscr{A}$ with respect to the origin.) Then (for $\mathscr{A},\mathscr{B},\mathscr{C} \in \mathbb{C}^n$, $\alpha \in \mathbb{C}$) the following is true:

(a) $$I(\mathscr{A}+\mathscr{B},\mathscr{C}) = I(\mathscr{A},\mathscr{C}) + I(\mathscr{B},\mathscr{C})$$

and

$$I(\mathscr{A},\mathscr{B}+\mathscr{C}) = I(\mathscr{A},\mathscr{B}) + I(\mathscr{A},\mathscr{C}),$$

(b) $$\alpha I(\mathscr{A},\mathscr{B}) = I(\alpha\mathscr{A},\mathscr{B}) = I(\mathscr{A},\bar{\alpha}\mathscr{B}),$$

(c) $$I(\mathscr{B},\mathscr{A}) = \overline{I(\mathscr{A},\mathscr{B})}\,.$$

When $I$ is replaced in the equations above by an arbitrary form $P\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}^n$, properties (a) and (b) define *sesquilinear* (or *conjugate bilinear*) forms, while (a), (b) and (c) define *Hermitian* forms (see, for example, [**14**, Chapter XI, §1]). There is another simple property of $I$ worth mentioning, namely

(d) $$I(x\mathscr{A}, x\mathscr{B}) = I(\mathscr{A},\mathscr{B}).$$

By a standard argument (see [**14**, p. 309] for example), for every sesquilinear form there is an $n \times n$ matrix $(h_{jk})$ such that $P(\mathscr{A},\mathscr{B}) = \Sigma_{j,k} h_{jk} a_j \bar{b}_k$. Property (c) forces $(h_{jk})$ to be a *Hermitian matrix*, that is $(h_{jk})^* = (h_{jk})$, where $(h_{jk})^*$ is defined to be $(\bar{h}_{kj})$. Property (d) is equivalent to the condition that $(h_{jk})$ be a circulant matrix; this property can be restated in terms of the polynomial $p(x) = \Sigma_{j<n} c_j x^j$, where $c_j = h_{j0}$, as the following theorem (whose proof is straightforward).

(3.1) *A form $P$ satisfies* (a), (b) *and* (d) *if and only if there exists a polynomial $p(x)$ for which $P(\mathscr{A},\mathscr{B}) = I(p(x)\mathscr{A},\mathscr{B})$ for all $n$-gons $\mathscr{A}$ and $\mathscr{B}$.*

Defining $(\Sigma_{j<n} c_j x^j)^* = \Sigma_{j<n} \bar{c}_j x^{n-j}$, we see that for all $\mathscr{A}$ and $\mathscr{B}$, $I(p(x)\mathscr{A},\mathscr{B}) = I(\mathscr{A}, p^*(x)\mathscr{B})$; hence *a form $I(p(x)$—, —$)$ determined by a polynomial $p(x)$ is Hermitian if and only if $p(x) = p^*(x)$.* We shall call forms satisfying (a), (b), (c) and (d) *polygonal forms*. These are the ones determined by polynomials $p(x)$ for which $p(x) = p^*(x)$ and include those of particular interest to the geometer. See [**9**], [**18**] and [**25**] for alternative approaches.

The quadratic form associated with a Hermitian (even sesquilinear) form $P(\mathscr{A},\mathscr{B})$ is defined by $P(\mathscr{A}) := P(\mathscr{A},\mathscr{A})$. It determines the Hermitian form since for all $\mathscr{A}$ and $\mathscr{B}$,

(3.2) $$2P(\mathscr{A},\mathscr{B}) = (P(\mathscr{A}+\mathscr{B}) - P(\mathscr{A}) - P(\mathscr{B})) + i(P(\mathscr{A}+i\mathscr{B}) - P(\mathscr{A}) - P(\mathscr{B}))$$

(see [**9**, p. 308]). We now introduce the form $F$(—), with $F$ standing for the German *Fläche*, that assigns area to an $n$-gon; we use (3.2) to pass to the associated polygonal form $F$(—, —). $F(\mathscr{A})$

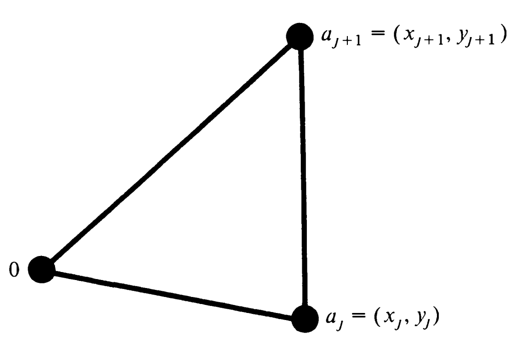



FIG. 3A. Area $= \frac{1}{2}(x_j x_{j+1} - x_{j+1} y_j) = (1/4i)(a_{j+1}\bar{a}_j - a_j \bar{a}_{j+1})$.

will assume both positive and negative values and thus represent a signed area. This is in contrast to $I(\mathscr{A})$, which is positive definite (i.e., $I(\mathscr{A}) > 0$ unless $\mathscr{A} = O$).

The area $F(\mathscr{A})$ of the $n$-gon $\mathscr{A}$ is defined to be the sum of the signed areas of the triangles $0a_j a_{j+1}$ (as in Fig. 3A):

$$F := F(\mathscr{A}) := \frac{1}{4i} \sum_{j<n} (a_{j+1}\bar{a}_j - a_j\bar{a}_{j+1});$$

from this it follows by an application of (3.2) that

$$F(\mathscr{A},\mathscr{B}) = \frac{1}{4i} \sum_{j<n} (a_{j+1}\bar{b}_j - a_j\bar{b}_{j+1}).$$

Finally (abbreviating as earlier $x^{n-1}$ by $x^{-1}$), we have that

$$F(\mathscr{A},\mathscr{B}) = \frac{1}{4i}[I(x\mathscr{A},\mathscr{B}) - I(\mathscr{A},x\mathscr{B})] = I\left(\frac{x - x^{-1}}{4i}\mathscr{A},\mathscr{B}\right), \tag{3.3}$$

which establishes that *signed area $F$ is the quadratic form of a polygonal form determined by the polynomial* $(x - x^{-1})/4i$.

The third and final form we shall consider here is $S = S(\mathscr{A})$, the sum of the squares of the side lengths of the polygon $\mathscr{A}$. We shall use it later in linking area and perimeter. As above, we can calculate the corresponding form $S(\mathscr{A},\mathscr{B})$ by using (3.2). Thus from the definition

$$S(\mathscr{A}) = \sum |a_{j+1} - a_j|^2 = \sum (a_{j+1} - a_j)(\bar{a}_{j+1} - \bar{a}_j),$$

we obtain that

$$S(\mathscr{A},\mathscr{B}) = \sum (a_{j+1} - a_j)(\bar{b}_{j+1} - \bar{b}_j),$$

and hence that

$$\begin{aligned} S(\mathscr{A},\mathscr{B}) &= I((x-1)\mathscr{A},(x-1)\mathscr{B}) = I((1-x)\mathscr{A},(1-x)\mathscr{B}) \\ &= I\big((2-(x+x^{-1}))\mathscr{A},\mathscr{B}\big). \end{aligned} \tag{3.4}$$

Therefore *the mapping assigning the sum of the squares of the sides to a polygon is a quadratic form arising from a polygonal form determined by the polynomial* $2 - (x + x^{-1})$.

Note that $S$ is positive definite on the subspace of $\mathbb{C}^n$ comprising the $n$-gons whose centroid is 0 (but it is only positive semi-definite on the whole space $\mathbb{C}^n$).

As an illustration of the use of these definitions we shall prove (cf. [**18**, (8.4)], [**25**, Th. 4], [**26**]):

(3.5) *If $\mathscr{A}$ is a k-affinely regular n-gon with centroid* 0, *then*

$$S(\mathscr{A}) = (2 - c_k)I(\mathscr{A}),$$

*where, as always,* $c_k = 2\cos(2k\pi/n)$.

*Proof.* Let 0 be the centroid of $\mathscr{A}$. Because $(x + x^{-1})\mathscr{A} = c_k\mathscr{A}$ (from (2.4)) we have that

$$S(\mathscr{A}) = I\big((2-(x+x^{-1}))\mathscr{A},\mathscr{A}\big) = I((2-c_k)\mathscr{A},\mathscr{A}) = (2-c_k)I(\mathscr{A}). \square$$

The theorem below is fundamental to all that follows. It permits us to extend automatically the validity of many types of statements from regular $n$-gons to arbitrary $n$-gons.

(3.6) Basis Theorem. *The regular n-gons $\mathscr{R}_0, \mathscr{R}_1, \ldots, \mathscr{R}_{n-1}$ form an orthogonal basis with respect to any polygonal form.*

*Proof.* We already know (from (2.8)) that $\mathscr{R}_0, \ldots, \mathscr{R}_{n-1}$ form a basis. So we only need to prove, for any polynomial $p(x)$ and for distinct integers $j$ and $k$ between 0 and $n-1$, that $I(p(x)\mathscr{R}_j, \mathscr{R}_k) = 0$. Now $p(x)\mathscr{R}_j = p(w^j)\mathscr{R}_j$ because $x\mathscr{R}_j = w^j\mathscr{R}_j$ for each $j$ (from (2.2)); therefore

$$I(p(x)\mathscr{R}_j,\mathscr{R}_k) = p(w^j)I(\mathscr{R}_j,\mathscr{R}_k) = p(w^j)\sum_{l<n} w^{(j-k)l}.$$

Because the sum of the first $n$ powers of any nontrivial $n$th root of unity is 0, the last expression is 0 and the theorem follows. □

From this theorem we can conclude immediately that *for all n-gons* $\mathscr{A}=\sum_{k<n}\alpha_k\mathscr{R}_k$ *and* $\mathscr{B}=\sum_{k<n}\beta_k\mathscr{R}_k$ *and for any polygonal form P*,

$$P(\mathscr{A},\mathscr{B}) = \sum_{k<n}\alpha_k\bar{\beta}_k P(\mathscr{R}_k) \quad\text{and}\quad P(\mathscr{A}) = \sum_{k<n}|\alpha_k|^2 P(\mathscr{R}_k). \tag{3.7}$$

A more substantial application of the Basis Theorem is the following addition to the Napoleon-Barlotti Theorem (2.6). It is a well-known result (see [**7**, 3.38] for an alternative approach) that the area of an arbitrary triangle equals the sum of the areas of its Barlotti triangles (which we defined after (2.6)). More generally,

(3.8) THEOREM. *If* $\mathscr{A}$ *is a k-affinely regular n-gon with centroid at* 0 *and* $\mathscr{B}$ *and* $\mathscr{C}$ *are its Barlotti n-gons, then*

$$(2+c_k)P(\mathscr{A}) = P(\mathscr{B}) + P(\mathscr{C}), \tag{$3.8_k$}$$

*where P is any polygonal form.*

For example, $P$ could be the area, the sum of the squared sides, or the polar moment of inertia ($F$, $S$, or $I$).

*Proof.* First observe that ($3.8_k$) just says $0=0$ when $w^k=-1$ (and $c_k=-2$); so let us assume $w^k\neq -1$. Also note that $|w^k|=1$ and

$$|w^k+1|^2 = (w^k+1)(w^{-k}+1) = 2+c_k.$$

By (2.6) $\mathscr{B}$ and $\mathscr{C}$ are $k$- and $(n-k)$-regular respectively; hence, by the Basis Theorem, they are orthogonal. Therefore, recalling (2.7), namely

$$\mathscr{A} = \frac{1}{w^k+1}\mathscr{B} + \frac{w^k}{w^k+1}\mathscr{C},$$

we obtain

$$\begin{aligned} P(\mathscr{A}) &= \frac{1}{|w^k+1|^2}P(\mathscr{B}) + \frac{|w^k|^2}{|w^k+1|^2}P(\mathscr{C}) \\ &= \frac{1}{2+c_k}(P(\beta)+P(\mathscr{C})). \ \square \end{aligned}$$

The theorem above has a partial converse which extends [**13**, Theorem 3]. Its proof will follow from the next result.

(3.9) THEOREM. *For any polygonal form P which is positive definite on the subspace of n-gons centered at* 0,

$$P\left(\frac{x+1}{2}\mathscr{A}\right) \leqslant \frac{2+c_1}{4}P(\mathscr{A});$$

*equality holds if and only if* $\mathscr{A}$ *is affinely regular and convex.*

*Proof.* We represent $\mathscr{A}$ as $\mathscr{A}=\sum_{k<n}\alpha_k\mathscr{R}_k$. Since $\mathscr{A}$ has centroid 0, $\alpha_0=0$ (see 2.9). Recalling (2.2) (that $x\mathscr{R}_k = w^k\mathscr{R}_k$) and noting that $c_k = 2\cos(2k\pi/n)\leqslant c_1$ for $1\leqslant k<n$,

$$P\left(\frac{x+1}{2}\mathscr{A}\right) = \sum_{k<n}|\alpha_k|^2 P\left(\frac{x+1}{2}\mathscr{R}_k\right)$$

$$
\begin{aligned}
&= \sum_{k<n} |\alpha_k|^2 P\left(\frac{w^k+1}{2}\mathscr{R}_k\right) \\
&= \sum_{k<n} |\alpha_k|^2 \frac{|w^k+1|^2}{4} P(\mathscr{R}_k) \\
&= \sum_{k<n} |\alpha_k|^2 \frac{2+c_k}{4} P(\mathscr{R}_k) \\
&\leqslant \frac{2+c_1}{4} \sum_{k<n} |\alpha_k|^2 P(\mathscr{R}_k) \\
&= \frac{2+c_1}{4} P(\mathscr{A}).
\end{aligned}
$$

Our assumption on $P$ forces $P(\mathscr{R}_k) > 0$ for all $k \neq 0$, which explains the inequality in the manipulations above; as a further consequence, equality holds if and only if $\alpha_k = 0$ for every $k \neq 1$, $n-1$, i.e., if and only if $\mathscr{A}$ is affinely regular and convex. □

The inequality in (3.9) is a comparison of the values of a form applied to an $n$-gon $\mathscr{A}$ and to its midpoint figure $\frac{1}{2}(x+1)\mathscr{A}$. If we replace $\mathscr{A}$ by $(x-1)\mathscr{B}$, we have an upper bound for $P(x^2-1)\mathscr{B}$, namely

$$P\big((x^2-1)\mathscr{B}\big) \leqslant (2+c_1)P\big((x-1)\mathscr{B}\big),$$

which is attained only for affinely regular convex $n$-gons. For the form $I$ this result coincides with Theorem 2 in [**13**], which compares the sum of the squared sides of an $n$-gon $\mathscr{B}$ with the sum of the squares of the diagonals from each $b_j$ to $b_{j+2}$.

We are now ready for the promised converse to (3.8).

*Let $\mathscr{A}$ be an $n$-gon that has its centroid at* 0, *$\mathscr{B}$ and $\mathscr{C}$ the outer and inner Barlotti $n$-gons of $\mathscr{A}$, and $P$ any polygonal form that is positive definite on the subspace of $n$-gons with centroid* 0. *If $(2+c_1)P(\mathscr{A}) = P(\mathscr{B}) + P(\mathscr{C})$, then $\mathscr{A}$ is convex and affinely regular.*

*Proof*. We substitute the formulas from Theorem (2.6),

$$\mathscr{B} = \frac{x-w^{-1}}{1-w^{-1}}\mathscr{A} \quad \text{and} \quad \mathscr{C} = \frac{x-w}{1-w}\mathscr{A},$$

into the given equation $(3.8_1)$ to obtain

$$
\begin{aligned}
(2+c_1)P(\mathscr{A}) &= P\left(\frac{x-w^{-1}}{1-w^{-1}}\mathscr{A}\right) + P\left(\frac{x-w}{1-w}\mathscr{A}\right) \\
&= \frac{1}{2-c_1}\Big(P\big((x-w^{-1})\mathscr{A}\big) + P\big((x-w)\mathscr{A}\big)\Big),
\end{aligned}
$$

so that

$$\left(4-c_1^2\right)P(\mathscr{A}) = 4P(\mathscr{A}) - c_1 P(x\mathscr{A},\mathscr{A}) - c_1 P(\mathscr{A}, x\mathscr{A}).$$

Therefore

$$
\begin{aligned}
c_1 P(\mathscr{A}) &= P(x\mathscr{A},\mathscr{A}) + P(\mathscr{A}, x\mathscr{A}) \\
&= P\big((x+1)\mathscr{A}\big) - 2P(\mathscr{A}),
\end{aligned}
$$

and finally,

$$(2+c_1)P(\mathscr{A}) = P\big((x+1)\mathscr{A}\big).$$

The result now follows immediately from (3.9).

**4. Isoperimetric Inequalities.** In this section we will apply Hermitian forms to obtain various isoperimetric inequalities. We have designed this section to be as independent of the first part of the paper as is possible. It depends only on the Basis Theorem (3.6) and the fact that signed area $F = F(\mathscr{A})$ and sum of squared sides $S = S(\mathscr{A})$ of an $n$-gon $\mathscr{A}$ are quadratic forms arising from Hermitian forms. We first establish the following statement.

(4.1) *For any n-gon* $\mathscr{A}$, $S \geqslant 4\tan(\pi/n)|F|$, *and equality holds if and only if* $\mathscr{A}$ *is regular and convex.*

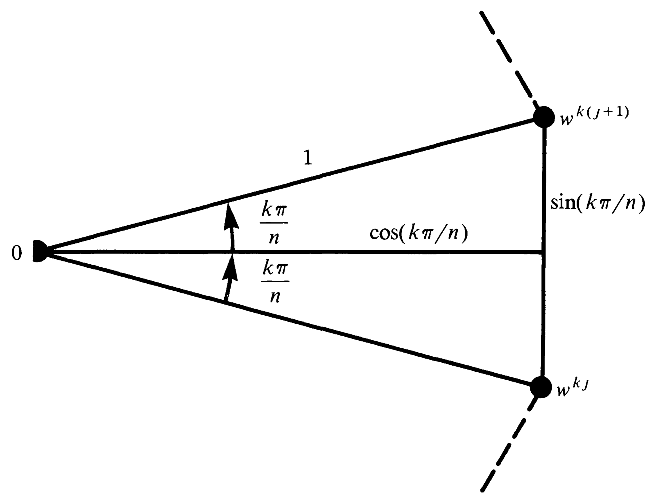



FIG. 4A. $F(\mathscr{R}_k) = n\sin(k\pi/n)\cos(k\pi/n)$; $S(\mathscr{R}_k) = 4n\sin^2(k\pi/n)$.

*Proof.* For each $k$ the $n$-gon $\mathscr{R}_k$ can be subdivided into $n$ congruent triangles each of which has a base of $2\sin(k\pi/n)$ and a height of $\cos(k\pi/n)$ (see Fig. 4A). So

$$F(\mathscr{R}_k) = n\sin(k\pi/n)\cos(k\pi/n)$$

and

$$S(\mathscr{R}_k) = 4n\sin^2(k\pi/n).$$

Because $|\tan(k\pi/n)| \geqslant \tan(\pi/n)$ for $k = 1, 2, \ldots, n-1$,

$$\begin{aligned} S(\mathscr{R}_k) &= 4n\sin^2(k\pi/n) \\ &= 4|\tan(k\pi/n)\,F(\mathscr{R}_k)| \\ &\geqslant 4\tan(\pi/n)\,|F(\mathscr{R}_k)|. \end{aligned}$$

The required inequality now follows from the Basis Theorem (3.6) since, for $\mathscr{A} = \sum_{k<n}\alpha_k\mathscr{R}_k$,

$$\begin{aligned} S = S\Big(\sum_{k<n}\alpha_k\mathscr{R}_k\Big) &= \sum_{k<n}|\alpha_k|^2 S(\mathscr{R}_k) \\ &\geqslant 4\tan(\pi/n)\Big|\sum_{k<n}|\alpha_k|^2 F(\mathscr{R}_k)\Big| \\ &= 4\tan(\pi/n)\Big|F\Big(\sum_{k<n}\alpha_k\mathscr{R}_k\Big)\Big| \\ &= 4\tan(\pi/n)\,|F|. \end{aligned}$$

An inspection shows that equality holds if and only if either all $\alpha_k = 0$ except possibly $\alpha_1$, or all $\alpha_k = 0$ except possibly $\alpha_{n-1}$, i.e., $\mathscr{A}$ is both regular and convex. □

There is another proof of this inequality which brings out the essential properties of the polynomials $p_S(x) = 2 - (x + x^{-1})$ and $p_F(x) = (x - x^{-1})/4i$ discussed in Section 3. Note that

$$(x+1)p_S(x) = -4i(x-1)p_F(x).$$

The function

$$f(\theta) = \left|\frac{p_S(e^{i\theta})}{p_F(e^{i\theta})}\right| = 4\left|\frac{e^{i\theta}-1}{e^{i\theta}+1}\right|$$

increases for $0 \leqslant \theta \leqslant \pi$ and decreases for $\pi \leqslant \theta \leqslant 2\pi$. Therefore its minimum value among $\theta = 2\pi/n, 4\pi/n, \dots, 2(n-1)\pi/n$ occurs for $\theta = \pm 2\pi/n$. Recalling that $p(x)\mathscr{R}_k = p(w^k)\mathscr{R}_k$ for any polynomial $p(x)$ (see the proof of (3.6)), we have that

$$S(\mathscr{R}_k) = p_S(w^k)I(\mathscr{R}_k) = \left|\frac{p_S(w^k)}{p_F(w^k)}\right| |p_F(w^k)| I(\mathscr{R}_k)$$

$$\geqslant \left|\frac{p_S(w)}{p_F(w)}\right| |p_F(w^k)| I(\mathscr{R}_k) = 4\tan(\pi/n)|F(\mathscr{R}_k)|.$$

The remainder of the proof is as above.

Now we can prove our next result.

(4.2) *For an n-gon $\mathscr{A}$ with perimeter L, $L^2 \geqslant 4n\tan(\pi/n)|F|$, and equality holds if and only if $\mathscr{A}$ is regular and convex.*

*Outline of the proof*. When $\mathscr{A}$ is equilateral $L^2 = nS$. (Just let the common edge be $e$ so that $L = ne$ and $S = ne^2$.) Thus for equilateral $n$-gons our claim is an immediate consequence of (4.1). The argument up to here is much like that of Schoenberg in [**25**], who attributes the result for equilateral $n$-gons to Jacob Steiner. For ways to extend the result to arbitrary $n$-gons we refer the reader to [**17**, Chapter 2], [**21**, Chapter IV], or [**22**, Chapter X]. An argument of a topological nature is common to all known proofs of (4.2) and appears to be inescapable (see [**17**, Section 2.4]). We shall skip the details.

Let us instead improve (4.1) in another direction by obtaining a more precise bound on the difference $S - 4\tan(\pi/n)F$. The first step is to prove for $k = 2, 3, \dots, n-1$ that

$$S(\mathscr{R}_k) - 4\tan(\pi/n)F(\mathscr{R}_k) \geqslant 8n\sin^2(\pi/n).$$

To this end we apply the formulas $S(\mathscr{R}_k) = 4n\sin^2(\pi/n)$ and $F(\mathscr{R}_k) = n\sin(k\pi/n)$ obtained above as well as some trigonometric manipulations and see that

$$\begin{aligned} S(\mathscr{R}_k) - 4\tan(\pi/n)F(\mathscr{R}_k) &= 4n\sin^2(k\pi/n) - 4n\tan(\pi/n)\sin(k\pi/n)\cos(k\pi/n) \\ &\geqslant 4n\sin^2(2\pi/n) - 4n\tan(\pi/n)\sin(2\pi/n)\cos(2\pi/n) \\ &= 8n\sin^2(\pi/n). \end{aligned}$$

The inequality above is easily checked for $2 \leqslant k \leqslant n-2$; when $k = n-1$, equality holds.

Next, because $S(\mathscr{R}_k) = 4\tan(\pi/n)F(\mathscr{R}_k)$ for $k = 0, 1$ it follows that, for $\mathscr{A} = \sum_{k<n}\alpha_k\mathscr{R}_k$,

$$\begin{aligned} S(\mathscr{A}) - 4\tan(\pi/n)F(\mathscr{A}) &= \sum_{k=2}^{n-1} |\alpha_k|^2 \big(S(\mathscr{R}_k) - 4\tan(\pi/n)F(\mathscr{R}_k)\big) \\ &\geqslant 8n\sin^2(\pi/n)\sum_{k=2}^{n-1}|\alpha_k|^2. \end{aligned}$$

We shall now estimate $\Sigma_{k=2}^{n-1}|\alpha_k|^2$ in terms of the vertices of $\mathscr{A}$. In order to express our conclusion most simply we shall assume that $\alpha_0 = 0$, i.e., that the centroid of $\mathscr{A}$ is 0 (see (2.9)). Moreover, also by (2.9), $\alpha_1 = \frac{1}{n}\Sigma_{j<n} a_j w^{-j}$ and so

$$|\alpha_1| \leqslant \frac{1}{n} \sum_{j<n} |a_j|.$$

We know that

$$I(\mathscr{A}) = \sum_{k<n} |\alpha_k|^2 I(\mathscr{R}_k) = \sum_{k<n} |\alpha_k|^2 \Big( \sum_{j<n} |w^{kj}|^2 \Big) = n \sum_{k<n} |\alpha_k|^2$$

and hence

$$\sum_{k<n} |\alpha_k|^2 = \frac{1}{n} \sum_{j<n} |a_j|^2.$$

Thus

$$\begin{aligned} \sum_{k=2}^{n-1} |\alpha_k|^2 &= \sum_{k<n} |\alpha_k|^2 - |\alpha_1|^2 \\ &\geqslant \frac{1}{n} \sum_{j<n} |a_j|^2 - \frac{1}{n^2} \Big( \sum_{j<n} |a_j| \Big)^2. \end{aligned} \tag{4.3}$$

The expression (4.3) is equal to $\sigma^2$, the variance of the set $\{|a_0|, |a_1|, \dots, |a_{n-1}|\}$; it measures the extent to which $\mathscr{A}$ differs from an $n$-gon inscribed in a circle whose radius is $\frac{1}{n}\Sigma_{j<n}|a_j|$. Combining the facts above we conclude for any $n$-gon $\mathscr{A}$ centered at 0 that

$$S(\mathscr{A}) - 4\tan(\pi/n) F(\mathscr{A}) \geqslant 8n\sigma^2 \sin^2(\pi/n).$$

As in (4.2) this inequality has an immediate consequence for equilateral $n$-gons.

(4.4) THEOREM. *For any equilateral n-gon $\mathscr{A}$ with centroid* 0, *perimeter $L$, and area $F$,*

$$L^2 - 4n\tan(\pi/n) F \geqslant 8n^2\sigma^2 \sin^2(\pi/n),$$

*where $\sigma^2$ is the variance of* $|a_0|, |a_1|, \dots, |a_{n-1}|$.

**5. Extension to Closed Curves.** The results of Section 4 extend naturally to a closed curve whose perimeter $L$ is well-defined. We shall denote the curve by $\mathscr{A}$ and its arc length parametrization by $a(s): [0, L] \to \mathbb{C}$, where $a(L) = a(0)$. We can consider the points of $a(s)$ as the limit points of the vertices $a_j$ of inscribed equilateral $n$-gons as $n \to \infty$. The finite Fourier series $a_j = \Sigma_{k<n}\alpha_k w^{jk}$ then becomes the Fourier series

$$a(s) = \sum_{k \in \mathbb{Z}} \alpha_k e^{2k\pi s i/L},$$

$$\text{where } \alpha_k = \frac{1}{L}\int_0^L a(s)\, e^{-2k\pi s i/L}\, ds.$$

Let us emphasize the obvious: A parameter other than $s$ will lead to a different Fourier series.

The quantities $L$, $F$, and $\sigma^2$ exist as integrals which are limits of the values for $n$-gons as $n$-increases to $\infty$. Explicitly

$$F := F(\mathscr{A}) := \frac{1}{2}\int_0^L \operatorname{Im} \bar{a}(s)\, a'(s)\, ds. \tag{5.1}$$

This formula can be recognized as Green's Theorem for areas expressed in the notation of complex numbers. The variance $\sigma^2$ of $|a(s)|$ is given by an integral obtained from (4.3), namely

$$\sigma^2 = \frac{1}{L}\int_0^L (|a(s)| - \mu)^2\,ds, \tag{5.2}$$

where

$$\mu = \frac{1}{L}\int_0^L |a(s)|\,ds.$$

The latter quantity $\mu$ can be thought of as the "average distance" of the curve from 0. When 0 is the centroid of the curve, $\mu$ is its "average radius".

(5.3) THEOREM. *Given a closed curve $\mathscr{A}$ with centroid* 0 *and perimeter $L$, which surrounds a region whose area is $F > 0$, then*

$$L^2 \geqslant 4\pi F + 8\pi^2\sigma^2.$$

*Moreover $L^2 = 4\pi F$ if and only if $\mathscr{A}$ is a circle.*

*Proof*. Inscribe an equilateral $n$-gon $\mathscr{A}_n$ in $\mathscr{A}$, or, to avoid a detour discussing whether this is always possible, follow W. Blaschke's approach, which consists in defining $\mathscr{A}_n$ as an equilateral $n$-gon whose vertices all lie on $\mathscr{A}$ except possibly for one which is "close" to $\mathscr{A}$ (see [**21**, p. 75]). Then (5.3) follows by taking the limit of the inequality in (4.4) for the polygons $\mathscr{A}_n$:

$$L_n^2 - 4n\tan(\pi/n)\,F_n \geqslant 8n^2\sigma_n^2\sin^2(\pi/n).$$

As $n \to \infty$, $L_n \to L$, $n\tan(\pi/n) \to \pi$, $F_n \to F$, $\sigma_n \to \sigma$, and $n^2\sin^2(\pi/n) \to \pi^2$. If $L = 4\pi F$, then $\sigma^2$ must equal 0 and so, by (5.2), $|a(s)| = \mu$ for every $s$. □

Theorem (5.3) is an example of what was termed in [**19**] and [**20**] a *Bonnesen-style* inequality—an isoperimetric inequality incorporating an error term that measures, in terms of some geometric property, the deviation of the curve from a circle. Our error term is not related to any of the results mentioned in [**20**]; a similar (but weaker) inequality was obtained in [**24**] by quite different methods.

Of course Theorem (5.3) can be proved directly, without recourse to our $n$-gon theory. This is the approach taken in a fascinating 1902 paper [**16**] of A. Hurwitz, the mentor of Hilbert and Minkowski. That paper contains among its many geometric applications of Fourier series a proof of the isoperimetric inequality that is close in spirit to ours (although Hurwitz failed to interpret the error term). His proof was reproduced in [**11**] and appears frequently in textbooks. Surprisingly, the Hurwitz paper contains virtually all of the published applications of Fourier series to geometry, even though it is clear that there is an untapped potential in his methods. Likewise, our $n$-gon theory demonstrates that there are many geometric applications of finite Fourier series as well; however, except for [**25**] their use has been restricted almost entirely to numerical techniques in such fields as applied harmonic analysis (see [**3**] or [**15**]). We hope that both our explanations and references provide the reader with some useful tools for geometry.

**Acknowledgements.** We would especially like to thank Adriano Barlotti for supplying us with reference [**4**]; Richard Guy for his helpful advice that included the reference that led us to Theorem (2.5); Murray Klamkin for his many ideas and references which greatly improved Section 4; Robert Osserman for reference [**24**]; and finally John Wilker who suggested replacing our original symmetric bilinear forms by the more elegant Hermitian forms in Section 3.

## References

**1.** F. Bachmann and E. Schmidt, $n$-gons (translation of $n$-Ecke, Bibliographisches Institut, Mannheim, 1970), Math. Expositions No. 18, Univ. of Toronto Press, 1975.

**2.** Adriano Barlotti, Una proprietà degli $n$-agoni che si ottengono trasformando in una affinità un $n$-agono regolare, Boll. Un. Mat. Ital. (3), 10 (1955) 96–98.

**3.** E. Oran Brigham, The Fast Fourier Transform, Prentice-Hall, Englewood Cliffs, NJ, 1974.

**4.** Vicenzo G. Cavallaro, Per la storia dei teoremi attribuiti a Napoleone Buonaparte e a Frank Morley, Archimede, 7–8 (1949) 286–287.

**5.** H. S. M. Coxeter, Affinely regular polygons, Abh. Math. Sem. Univ. Hamburg, 34 (1969) 38–58.
**6.** ______, Introduction to Geometry (2nd ed.), Wiley, New York, 1969.
**7.** H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Math. Library, Math. Assoc. Amer., Washington, DC, 1967.
**8.** Philip J. Davis, Cyclic transformations of polygons and the generalized inverse, Canad. J. Math., 29 (1977) 756–770.
**9.** ______, Circulant Matrices, Wiley, New York, 1979.
**10.** Clement V. Durell, Modern Geometry, MacMillan, London, 1934.
**11.** Richard Courant and David Hilbert, Methods of Mathematical Physics, vol. 1, translated and revised from the German original, Interscience, New York, 1961.
**12.** J. C. Fisher, D. Ruoff, and J. Shilleto, Polygons and polynomials, in The Geometric Vein: the Coxeter Festschrift, edited by C. Davis, B. Grünbaum, and F. A. Sherk, Springer, New York, 1981, pp. 321–333.
**13.** Leon Gerber, Napoleon's theorem and the parallelogram inequality for affine-regular polygons, this MONTHLY, 87 (1980) 644–648.
**14.** W H. Greub, Linear Algebra, 3rd ed., Springer, New York, 1967.
**15.** Richard W. Hamming, Introduction to Applied Numerical Analysis, McGraw-Hill, New York, 1971.
**16.** Adolf Hurwitz, Sur quelques applications géométriques des séries de Fourier, in Mathematische Werke, vol. 1, Basel, 1932, pp. 509–554.
**17.** Nicholas D. Kazarinoff, Geometric Inequalities, New Math. Library, Math. Assoc. Amer., Washington, DC, 1961.
**18.** B. H. Neumann, Some remarks on polygons, J. London Math. Soc., 16 (1941) 230–245.
**19.** Robert Osserman, The isoperimetric inequality, Bull. Amer. Math. Soc., 84 (1978) 1182–1238.
**20.** ______, Bonnesen-style isoperimetric inequalities, this MONTHLY, 86 (1979) 1–29.
**21.** D. Pedoe, Circles: A Mathematical View, Dover, New York, 1979.
**22.** G. Pólya, Induction and Analogy in Mathematics, Mathematics and Plausible Reasoning, vol. 1, Princeton Univ. Press, Princeton, 1954.
**23.** D. Ruoff and J. Shilleto, Recursive polygons, Boll. Un. Mat. Ital. (5), 15-B (1978) 968–981.
**24.** Horst Sachs, Ungleichungen für Umfang, Flächeninhalt und Trägheitsmoment konvexer Kurven, Acta. Math. Acad. Sci. Hungar., 11 (1960) 103–115.
**25.** I. J. Schoenberg, The finite Fourier series and elementary geometry, this MONTHLY, 57 (1950) 390–404.
**26.** O. Shisha, On the discrete version of Wirtinger's inequality, this MONTHLY, 80 (1973) 755–760.

# CRITICAL POINTS AND SIGMOIDICITY OF POSITIVE RATIONAL FUNCTIONS

W. G. BARDSLEY
*Department of Obstetrics and Gynaecology, University of Manchester, St. Mary's Hospital, Whitworth Park, Manchester, M13 0JH, England*

R. M. W. WOOD
*Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, England*

**1. Introduction.** The problem of sketching the graph of a rational function

$$r(x) = \frac{a_0 + a_1 x + \cdots + a_n x^n}{b_0 + b_1 x + \cdots + b_m x^m}$$

---

W. G. Bardsley was awarded the degree of Ph.D. (Chemistry) by the University of Manchester (U.K.) in 1963 for research into the structure, degradation and synthesis of indole alkaloids. In 1976 the University of Manchester awarded him the degree of D.Sc. (Biochemistry) for practical studies in enzymology and theoretical work on enzyme kinetics. His recreational interests include weightlifting, mountaineering, ornithology, four-wheel drive vehicles, mathematics, history of science, and music, especially organ playing.

R. M. W. Wood was awarded the degree of Ph.D. by the University of Oxford in 1966 for research in algebraic topology on the subject of Bott periodicity and Banach algebras.

is as old as calculus itself and textbooks abound with worked examples. The biochemist, however, is interested in a restricted form of the problem in which the coefficients of the polynomials

$$f(x) = a_0 + a_1x + \cdots + a_nx^n,$$
$$g(x) = b_0 + b_1x + \cdots + b_mx^m,$$

are required to be nonnegative and only the part of the graph of $r(x)$ in the positive quadrant is of significance. We shall not attempt to explain the numerous situations in life sciences where the study of such graphs is important nor the chemical reactions which give rise to the rational functions $r(x)$. Such information can be found elsewhere [**1**], [**2**], [**3**], [**4**]. Let it suffice to mention that, to the experimental biochemist, theoretical results concerning the shapes of rational functions are of considerable interest. So in this paper we collect together a number of mathematical theorems concerning the turning points, inflexions, and sigmoidicity of rational functions.

As an example, we quote the following result.

1.1. THEOREM. *The number of critical points in the positive quadrant of a nonconstant rational function*

$$\frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m},$$

*with all coefficients $a_i, b_j$ nonnegative, cannot exceed the minimum of the degrees $m, n$. If $m = n$, then this number cannot exceed $m - 1$.*

The proof of this theorem is taken up in Section 2, where we further show that the result is best possible. A necessary condition for achieving the maximum number of critical points is also indicated. Related problems concerning the distribution of critical points and inflexions are discussed in Section 3. Finally in Section 4 we define the notion of sigmoidicity for a certain class of rational functions and solve the problem of maximising sigmoidicity up to degree four.

**2. Critical Points.** Consider two polynomials

$$f(x) = a_0 + a_1x + \cdots + a_nx^n,$$
$$g(x) = b_0 + b_1x + \cdots + b_mx^m,$$

with $a_nb_m \neq 0$. The Wronskian is defined by

$$W(f,g) = fg' - f'g = -W(g,f),$$

where $f'$ and $g'$ are the derivatives of $f$ and $g$, respectively. We are particularly interested in the situation where all coefficients $a_i, b_j$ are nonnegative. The positive zeros of $W(f,g)$ are the positive critical points of the rational function $r = f/g$ because $r' = W(g,f)/g^2$. We begin, however, with a version of Rolle's theorem for suitably differentiable functions $f$ and $g$. Our main results follow as corollaries.

2.1. THEOREM. *Let $I = [a,b]$ be an interval of the real line and suppose that neither $g$ nor $g'$ has a zero in $I$. Then between two zeros of $W(f,g)$ in $I$ there is at least one zero of $W(f',g')$.*

To prove the theorem we first expand the identically zero determinant

$$\begin{vmatrix} g & g' & g'' \\ g & g' & g'' \\ f & f' & f'' \end{vmatrix} = 0,$$

to obtain

$$gW(g',f') = g'W'(g,f) - g''W(g,f).$$

An application of Rolle's theorem to $W(g,f)/g'$ then shows that between two zeros of $W(g,f)$

there is at least one zero of $W(g', f')$. The argument requires that $g, g'$ do not vanish in the interval.

This concludes the proof of Theorem 2.1. It should be noted of course that the roles of $f$ and $g$ may be interchanged because $W(f, g) = -W(g, f)$.

For our first corollary we assume that the coefficients of $g$ are all nonnegative. The same is then true for the derivatives of $g$. We take $I = [0, \infty)$ as our interval.

2.2. COROLLARY. *Let $f(x), g(x)$ be polynomials of degrees $n, m$, respectively, and suppose $g(x)$ has nonnegative coefficients. If $n < m$, then the number of positive zeros of $W(f, g)$ cannot exceed $n$. If $n = m$ then this number cannot exceed $n - 1$.*

Theorem 1.1 is an immediate consequence of this corollary. The proof of Corollary 2.2 is by induction on degrees, but first of all we make a few preliminary remarks about the nature of the result. The degree of $W(f, g)$ is $m + n - 1$ unless $m = n$ in which case it is $m + n - 2$ or less. Theorem 1.1 does therefore impose a significant restriction on the number of critical points $f/g$ in the positive quadrant. When enumerating zeros of polynomials we always take account of multiplicity. For example if $f(x) = x^n, g(x) = x^m, m \neq n$, then

$$W(f, g) = (m - n)x^{m+n-1},$$

which has $m + n - 1$ zeros at $x = 0$. Incidentally this does not contradict Corollary 2.2 which only refers to (strictly) positive zeros. Let us suppose then, contrary to the statement of the corollary, that $W(f, g)$ has $n + 1$ positive zeros in case $n < m$ or $n$ positive zeros in case $n = m$. Applying Theorem 2.1 iteratively $n$ times with $I = [0, \infty)$, we arrive at a situation where the new $f$ is constant say $f(x) \equiv c$ and the Wronskian $W(f, g) = cg'$ has at least one positive zero. This is a contradiction and proves Theorem 2.1 in case $n < m$. If $n = m$, we apply Theorem 2.1 iteratively $n - 1$ times to arrive at a situation where the new $f, g$ are linear expressions $ax + b, cx + d$, respectively. If at any stage of the argument the Wronskian becomes identically zero, say $W(f', g') \equiv 0$, then $f(x) = kg(x) + l$ for some constants $k$, $l$ and $W(f, g) = lg'$ and the conclusion follows. Otherwise at the last stage where $f(x) = ax + b$, $g(x) = cx + d$, the Wronskian $W(f, g) = bc - ad$ is supposed to have one positive zero, which is a contradiction.

A special case of a rational function of biological interest [**4**] with $\deg(f) = \deg(g) = n$ that can have no positive turning points is

$$r(x) = xF'/nF,$$

where $F(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$; $a_i \geqslant 0$.

Turning points occur at zeros of $FF' + x(FF'' - F'^2)$ and all coefficients of this polynomial are nonnegative since

$$FF' + x(FF'' - F'^2) = \sum_{i,j} ja_ia_j(j - i)x^{i+j-1}, \qquad 0 \leqslant i + j \leqslant 2n,$$

and so the coefficient of $x^{i+j-1}$ is $a_ia_j(j - i)^2$.

However, we now present examples to show that the maximum number of turning points permitted by Theorem 1.1 can be achieved. First of all let

$$f(x) = (x + 1)^n + (x - 1)^n, \; g(x) = (x + 1)^n.$$

In this case

$$W(f, g) = -2n(x + 1)^{n-1}(x - 1)^{n-1},$$

which has a zero of multiplicity $n - 1$ at $x = 1$. Next consider

$$g(x) = \frac{(x - 1)^m}{m!} + \frac{(x - 1)^{m-1}}{(m - 1)!} + \cdots + (x - 1) + 1,$$

the first $m + 1$ terms in the expansion of $e^{x-1}$. It can be checked that all coefficients of $g(x)$ are nonnegative when $g(x)$ is written out as a polynomial in $x$. Taking

$$f(x) = \frac{(x-1)^n}{n!} + \frac{(x-1)^{n-1}}{(n-1)!} + \cdots + (x-1) + 1,$$

we find that

$$W(f,g) = (x-1)^n \left\{ \frac{g}{n!} - \frac{(x-1)^{m-n}}{m!} f \right\},$$

which has a zero of multiplicity $n$ at $x = 1$.

A necessary condition for achieving the maximum number of turning points can be stated in terms of the expressions

$$C_i(f,g) = a_i b_{i-1} - a_{i-1} b_i.$$

We interpret $a_i = 0$ unless $0 \leqslant i \leqslant n$ and $b_j = 0$ unless $0 \leqslant j \leqslant m$.

2.3. PROPOSITION. *Let $f(x), g(x)$ be polynomials with nonnegative coefficients and suppose $W(f,g)$ has the maximum possible number of positive zeros for the given degrees of $f$ and $g$. Then the expressions $C_i(f,g)$ alternate in sign in the sense that $C_i C_{i+1} \leqslant 0$ for all $i$.*

To prove the result we note first of all that

$$C_k(f',g') = k(k+1) C_{k+1}(f,g),$$

which expresses $C_k(f,g)$ in terms of $C_{k-1}(f',g')$ except $k = 1$. If the proposition is true for $f', g'$, it follows immediately for $f, g$, except possibly for $C_1 C_2$. However, this case is also covered by making the substitution $x \mapsto 1/x$. To see how this works consider the case $m = n$. Take

$$p(x) = x^n f(1/x),$$
$$q(x) = x^n g(1/x),$$

whereupon $C_i(f,g) = -C_{n-i+1}(p,q)$. In particular,

$$C_1(f,g) C_2(f,g) = C_n(p,q) C_{n-1}(p,q).$$

If we assume that $W(f,g)$ has the maximum possible number of positive zeros, then the same is true of $W(p,q)$, $W(f',g')$ and $W(p',q')$. This establishes an inductive step and to start the induction we consider two cases. First of all let

$$f(x) = a_0 + a_1 x + a_2 x^2,$$
$$g(x) = b_0 + b_1 x + b_2 x^2,$$

and suppose $f/g$ has one positive turning point (the maximum permitted number according to Theorem 1.1). In this case

$$W(f,g) = -C_2 x^2 - 2(a_2 b_0 - a_0 b_2) x - C_1,$$

which has one positive zero and must therefore have a negative one. Hence

$$C_1 C_2 \leqslant 0,$$

with all other $C_i C_{i+1}$ equal to zero. In the second case we take

$$f(x) = a_0 + a_1 x,$$
$$g(x) = b_0 + b_1 x + \cdots + b_n x^n,$$

such that $f/g$ has a positive turning point. Then

$$W(f,g) = (n-1) a_1 b_n x^n + (n a_0 b_n + (n-2) a_1 b_{n-1}) x^{n-1} + \cdots$$
$$+ 2 a_0 b_2 x + (a_0 b_1 - a_1 b_0).$$

The coefficients here are all nonnegative except possibly

$$-C_1 = a_0 b_1 - a_1 b_0,$$

and this must be negative by Descartes' rule of signs because $W(f, g)$ is assumed to have a positive zero. Hence

$$C_1 > 0, \qquad C_2 = -a_1 b_2 < 0,$$

and therefore $C_1 C_2 < 0$. This completes the inductive argument and the proof of Proposition 2.3.

**3. Inflexion Points.** After turning points, the next most salient features of a graph are its inflexions.

3.1. THEOREM. *Let $f(x), g(x)$ be polynomials with nonnegative coefficients of degree $n, m$ respectively. If $n < m$, then the number of positive inflexions of $f/g$ cannot exceed $m + n - 1$. If $m = n$ or $m = n - 1$, then this number cannot exceed $2m - 2$. If $m < n - 1$, then the number of positive inflexions cannot exceed $2m$.*

We should note that the statement of Theorem 1.1 is symmetrical in $m, n$ because the positive turning points of $f/g$ and $g/f$ are the same. This is not true for inflexions. The positive inflexions of $f/g$ are the positive turning points of $W(f, g)/g^2$. Now $g^2$ has nonnegative coefficients. If $n < m$, then

$$\text{degree } W(f, g) \leqslant m + n - 1 < 2m = \text{degree } g^2.$$

We may, therefore, apply Corollary 2.2 with $W(f, g)$ replacing $f$, and $g^2$ replacing $g$, to prove the first part of Theorem 3.1. To deal with the cases $m = n$ or $m = n - 1$ we can always find constants $k, l$ such that the degree of the numerator in

$$f/g + kx + l = (f + (kx + l) g)/g$$

will be less than the degree of the denominator. The inflexion points are not changed by adding the linear expression $kx + l$. Now we apply the previous case. It remains to deal with the case $m < n - 1$ and for this we need a preliminary result.

3.2. LEMMA. *Let $f, g$ be polynomials with nonnegative coefficients of degrees $n, m$ respectively. If $n > m + 1$, then the $2m$th derivative of the Wronskian, $W^{(2m)}(f, g)$, has degree $n - m - 1$ and nonnegative coefficients.*

The proof relies on Leibniz's formula for iterated differentiation of a product, applied to the Wronskian $W(g, f) = gf' - g'f$, namely

$$W^{(2m)}(g, f) = \sum_{i=0}^{2m} \binom{2m}{i} g^{(i)} f^{(2m-i+1)} - \sum_{i=0}^{2m} \binom{2m}{i} f^{(i)} g^{(2m-i+1)}.$$

Now $g^{(i)}$ vanishes for $i > m$ since $g$ is a polynomial of degree $m$. Hence, with $i - 1$ in place of $i$ in the second summation, we have

$$\begin{aligned} W^{(2m)}(g, f) &= \sum_{i=0}^{m} \binom{2m}{i} g^{(i)} f^{(2m-i+1)} - \sum_{i=1}^{m} \binom{2m}{i-1} g^{(i)} f^{(2m-i+1)} \\ &= g f^{(2m+1)} + \sum_{i=1}^{m} \left[ \binom{2m}{i} - \binom{2m}{i-1} \right] g^{(i)} f^{(2m-i+1)}. \end{aligned}$$

In the range $1 \leqslant i \leqslant m$, the difference of binomial coefficients, $\binom{2m}{i} - \binom{2m}{i-1}$, is positive. This completes the proof of the lemma.

We return to the final part of Theorem 3.1. Let us assume, contrary to the statement of the theorem, that $n > m + 1$ and $f/g$ has $2m + 1$ positive inflexions. Applying Theorem 2.1 iteratively $2m$ times with $I = [0, \infty)$, $W(f, g)$ in place of $f$, and $g^2$ in place of $g$, we arrive at a situation where the Wronskian of the functions $W^{(2m)}(f, g)$ and $(g^2)^{(2m)}$ is supposed to have a

positive zero. But $(g^2)^{(2m)}$ is a nonzero constant $k$, say, and the Wronskian is then $kW^{(2m+1)}$. This has nonnegative coefficients by Lemma 3.2 and we have a contradiction, which completes the proof of Theorem 3.1.

To make further progress with shape-classification for graphs of positive rational functions in the positive quadrant, we might ask about the relative distribution of turning points and inflexions. A simple example to consider is

$$t(x) = \frac{(x-1)(x-2)\cdots(x-n)}{(1+x)^n} + K,$$

where $K$ is a sufficiently large positive constant to lead to nonzero numerator coefficients. It can be checked that $t(x)$ has $n-1$ turning points and $n-1$ inflexions in the positive quadrant which interlace.

Using the abbreviation $T$ for turning point and $I$ for inflexion, and reading the graph of $t(x)$ in the positive quadrant from left to right, we have a regular pattern

$$\ldots TITI \ldots$$

of turning points and inflexions which we may call the "symbol" of the graph. A primitive form of shape classification would entail the listing of all possible symbols for rational functions with positive coefficients of given degree $n$ over $m$. An analysis of low degree cases, reveals, for example, that the symbol

$$I^2 TITI$$

is realisable by cubics over cubics [**9**]. Here the maximum numbers of inflexions and turning points permitted by Theorems 1.1 and 3.1 are simultaneously realisable. For higher degrees the symbolic classification remains a problem. It can be solved in certain simple cases. For example, the reciprocal $1/g(x)$ of a polynomial $g(x)$ with positive coefficients cannot have any positive turning points but it is possible to find $g(x)$ of degree $m$ such that $1/g(x)$ has $m-1$ positive inflexions. In other words the symbol $I^{m-1}$ is realisable. This symbol is maximal and by reading the graph of $1/g(x)$ from suitable points to the right of the origin we can then realise symbols $I^r$ for $0 \leqslant r < m$.

A special case of biochemical interest is a rational function of the form

$$y = \sum_{i=1}^{n} y_i,$$

where $y_i = n_i K_i x/(1 + K_i x)$, $n_i, K_i \geqslant 0$. Clearly $y' > 0$ and $y'' < 0$ for $x > 0$, and so the $y(x)$ curve has no turning points or inflexions. Biochemists also use transformation of axes and a popular plot is in $(1/x, 1/y)$ space. There are no inflexions of $y(x)$ in this space either and to see this we first note that

$$d^2(1/y)\,d(1/x)^2 < 0 \Leftrightarrow 2y'(xy' - y) - xyy'' < 0.$$

To establish the result by induction we suppose that the graph of $1/y$ against $1/x$ is strictly concave down for $n = k-1$, which is known to be the case for $n = 2$ and 3 [**1**]. Then we consider the effect of adding a further term $y_k$.

Now if $2y'(xy' - y) - xyy'' < 0$ for $y = y_1 + y_2 + \cdots + y_{k-1}$, the sign after adding $y_k$ will depend upon $P + Q$, where

$$P = 2y'(xy' - y) - xyy'' + 2y_k'(xy_k' - y_k) - xy_k y_k''$$

and

$$Q = 2y_k'(2xy' - y) - y_k(2y' + xy'') - xyy_k''.$$

As $P < 0$ we only need to consider a typical term, $q_{jk}$, of $Q = \sum_{j=1}^{k-1} q_{jk}$ resulting from $y_j$ and $y_k$. Since

$$q_{jk} = \frac{-2n_j n_k K_j K_k (K_j - K_k)^2 x^3}{(1 + K_j x)^3 (1 + K_k x)^3} \leqslant 0,$$

$Q \leqslant 0$ and the desired result is established.

To conclude this section we examine one important example which has received attention in the biochemical literature [4]. We first point out that the Hessian of a polynomial $p(x)$ is defined by

$$H(p) = npp'' - (n-1)(p')^2,$$

where $n =$ degree $(p)$. Taking

$$f(x) = np(x) - xp'(x),$$
$$g(x) = p'(x),$$

we find that

$$H(p) = W(f, g).$$

Theorem 2.1 then yields the following result.

3.3. PROPOSITION. *Let $p(x)$ be a polynomial of degree $n$ with nonnegative coefficients and suppose the Hessian is not identically zero i.e., $p(x) \neq (ax + b)^n$. Then $H(p)$ has at most $n - 2$ positive zeros.*

Polynomials, $p(x)$, of the form

$$p(x) = 1 + \binom{n}{1} K_1 x + \binom{n}{2} K_1 K_2 x^2 + \cdots + K_1 K_2 \cdots K_n x^n, \qquad K_i \geqslant 0,$$

have importance in biochemistry and the coefficients

$$\gamma_i = \log(K_{i+1}/K_i),$$

known as cooperativity coefficients, have attracted much attention [4]. When $H(p)$ has the maximum number of positive zeros, then the positive, rational function $(np(x) - xp'(x))/p'(x)$ has the maximum number of positive critical points. However, for this rational function we find that

$$C_i = i(n-i)\binom{n}{i}^2 (K_1 K_2 \cdots K_{i-1})^2 K_i (K_i - K_{i+1}),$$

and so the sign of $C_i C_{i+1}$ is the same as the sign of $\gamma_i \gamma_{i+1}$. Our previous result (2.3) shows that if $H(p)$ has the maximum number of positive zeros, then

$$\gamma_i \gamma_{i+1} \leqslant 0.$$

Using a theorem of Newton that was established by Sylvester, discussed in a textbook by Todhunter [8] but largely overlooked since [5], we can make further conclusions about the nature of the zeros of $p(x)$ when $\gamma_i \gamma_{i+1} \leqslant 0$. If the sequence $\log(K_2/K_1), \log(K_3/K_2), \ldots, \log(K_n/K_{n-1})$ has alternating signs $- + - \cdots \pm$, then $p(x)$ has at least $n - 2$ nonreal zeros if $n$ is even, and exactly $n - 1$ nonreal zeros if $n$ is odd. If the signs alternate in the order $+ - + \cdots \pm$, then there are still $n - 1$ nonreal zeros if $n$ is odd, but $n$ nonreal zeros if $n$ is even.

**4. Sigmoidicity.** In this section we consider rational functions of the form $r(x) = xv(x)$, where

$$v(x) = \frac{a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0},$$

with $n > 1$, $a_i$ and $b_j \geqslant 0$ for $0 \leqslant i \leqslant n-1$ and $0 \leqslant j \leqslant n$, $b_0 b_n \neq 0$. We are looking therefore at

rational functions of degree $n$ over $n$ whose coefficients are nonnegative and whose graphs go through the origin. We further assume that the graph of $r(x)$ is $S$-shaped or "sigmoid" as it issues from the origin, as in Fig. 1:

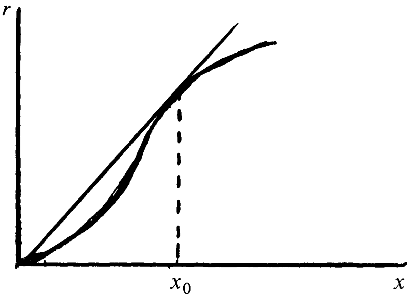



FIG. 1

The sigmoidicity of such curves has been regarded as of great importance to the biological function and numerous attempts to explain and quantify this phenomenon have been made [**2**], [**3**], [**6**], [**7**].

Sigmoidicity occurs if $r''(0) \geqslant 0$, which is equivalent to $v'(0) \geqslant 0$. Let $x_0$ be the first positive value of $x$ for which a line through the origin meets the graph of $r(x)$ tangentially as illustrated above. Equivalently $x_0$ is the first positive turning point of $v(x)$.

The "sigmoidicity" of $r(x)$ can be measured by the ratio of the area of the triangle with vertices at $(0,0),(x_0,0),(x_0,r(x_0))$ to the area under the graph of $r(x)$ for $x$ between 0 and $x_0$. This works out to be

$$\frac{x_0 r(x_0)}{2\int_0^{x_0} r(x)\,dx} = \frac{1}{2\int_0^{x_0} \dfrac{xv(x)}{x_0^2 v(x_0)}\,dx}.$$

The problem is to maximise sigmoidicity over all $r(x)$ of given degree $n$ over $n$ subject to the prescribed conditions. The expression above for sigmoidicity is invariant under rescaling of $x$ and $r(x)$, i.e., multiplication of these quantities by positive constants. It is therefore convenient to normalise the problem by introducing the function

$$u(x) = \frac{v(xx_0)}{v(x_0)}.$$

Then $u(1) = 1, u'(1) = 0$ and the problem boils down to minimising the integral

$$\int_0^1 xu(x)\,dx$$

over all

$$u(x) = \frac{a_{n-1}x^{n-1} + \cdots + a_1 x + a_0}{b_n x^n + \cdots + b_1 x + b_0},$$

for given $n > 1$ subject to the conditions that all $a_i, b_j$ are nonnegative, $b_0 b_n \neq 0$, $u'(0) \geqslant 0$, $u(1) = 1$ and $x = 1$ is the first zero of $u'(x)$.

The first step is to establish a result about pointwise minimisation of functions $u(x)$ in the interval $0 \leqslant x \leqslant 1$. Let $C$ denote the class of functions $u(x)$ as above and note that the conditions $u(1) = 1, u'(1) = 0$ are equivalent to

$$a_{n-1} + a_{n-2} + \cdots + a_1 + a_0 = b_n + b_{n-1} + \cdots + b_1 + b_0,$$
$$(n-1)a_{n-1} + (n-2)a_{n-2} + \cdots + a_1 = nb_n + (n-1)b_{n-1} + \cdots + b_1.$$

For each $k, n-1 \leqslant k \leqslant 1$, the following rational function is a member of $C$:

$$w(x)=\frac{[(n-k+1)-(k-1)\beta]x^k+[k-n+k\beta]x^{k-1}}{x^n+\beta},$$

providing $\beta$ lies in the range

$$\frac{n-k}{k}\leqslant\beta\leqslant\frac{n-k+1}{k-1}$$

for $k>1$ and $\beta\geqslant n-1$ for $k=1$. This guarantees the positivity of the coefficients, and the conditions $w(1)=1, w'(1)=0$ are satisfied because

$$n-k+1-(k-1)\beta+k-n+k\beta=1+\beta,$$
$$k[n-k+1-(k-1)\beta]+(k-1)[k-n+k\beta]=n.$$

Thus for each $k$ and each parameter $\beta$ in the stipulated range we have a $w(x)$ in the class $C$. The special nature of $w(x)$ is that in the numerator the only nonzero coefficients belong to adjacent powers of $x$ and in the denominator the only nonzero coefficients belong to the constant term and highest power of $x$.

4.1. PROPOSITION. *For each member $u(x)$ of the class $C$ there is at least one $w(x)$ such that $w(x)$ pointwise minimises $u(x)$ for all $x\geqslant 0$.*

This means that $w(x)\leqslant u(x)$ for $x\geqslant 0$. The condition that $x=1$ is the first zero of $u'(x)$ is irrelevant in this proposition. We may take $x=1$ to be any turning point of $u(x)$. The idea of the proof of Proposition 4.1 is first of all to reduce the numerator of $u(x)$ keeping the denominator fixed and then to increase the denominator keeping the numerator fixed. Consider then three numerator coefficients

$$a_k, a_{l+1}, a_l,$$

for $k>l+1$. We can replace them by

$$a_k'=a_k+\lambda, a_{l+1}'=a_{l+1}+\mu, a_l'=a_l+\nu,$$

without changing the conditions $u(1)=1, u'(1)=0$ providing

$$\lambda+\mu+\nu=0,$$
$$k\lambda+(l+1)\mu+l\nu=0.$$

This gives

$$\lambda=\frac{\nu}{k-l-1},$$
$$\mu=\frac{-(k-l)\nu}{k-l-1}.$$

The change in the numerator of $u$ obtained by replacing $a_k, a_{l+1}, a_l$ by $a_k', a_{l+1}', a_l'$ is

$$e(x)=\frac{\nu x^l}{k-l-1}\left[x^{k-l}-(k-l)x+k-l-1\right].$$

Now the derivative of the function

$$d(x)=x^{k-l}-(k-l)x+k-l-1$$

is given by

$$d'(x)=(k-l)(x^{k-l-1}-1),$$

with second derivative

$$d''(x)=(k-l)(k-l-1)x^{k-l-2}.$$

It follows that $d(x) \geqslant 0$ for all $x \geqslant 0$ with minimum value equal to 0 at $x = 1$. Hence by choosing $\nu$ negative we ensure that $e(x) \leqslant 0$ for all $x \geqslant 0$ and the numerator of $u(x)$ is thereby pointwise decreased. The formulae above for $a_k', a_{l+1}', a_l', \lambda, \mu$ show that the effect of taking $\nu$ negative is to decrease $a_k$ and $a_l$ but to increase $a_{l+1}$. Consequently we can eliminate either $a_k$ or $a_{l+1}$, depending on which is the smaller, whilst decreasing $u$, preserving the nonnegativity of all coefficients and, of course, remaining in the class $C$. Starting with the triple $a_{n-1}, a_1, a_0$ and working inwards, we can eliminate all coefficients $a_i$ except possibly certain two adjacent ones whilst decreasing the numerator of $u(x)$ and therefore $u(x)$ itself.

Now we carry out a similar programme on the denominator of $u(x)$ which of course we wish to increase. Using an analogous notation with the denominator coefficients $b_i$ replacing the numerator coefficients $a_i$, we see that, by taking $\nu$ positive, we increase $b_k$ and $b_l$ but decrease $b_{l+1}$ whilst increasing the denominator of $u(x)$. This time we can eliminate the middle term of any triple of coefficients, preserving positivity of the rest. Starting with the triple $b_n, b_{n-1}, b_{n-2}$ and working to the right, we eliminate all denominator coefficients except $b_n$ and $b_0$. This leads finally to a member of $C$ of the form

$$w(x) = \frac{a_k' x^k + a_{k-1}' x^{k-1}}{b_n' x^n + b_0'},$$

pointwise minimising $u(x)$ for all $x \geqslant 0$. We get the form of $w(x)$ stated above using the conditions $w(1) = 1, w'(1) = 0$ to work out $a_k', a_{k-1}'$ in terms of $\beta = b_0'/b_n'$. This completes the proof of Proposition 4.1.

As far as the original problem of minimising $\int_0^1 xu(x)\,dx$ is concerned, we know now that we may restrict attention to the special functions $u(x) = w(x)$. There is a 1-parameter family of such $w(x)$ for each $k, n-1 \leqslant k \leqslant 1$. How much further can we go in pointwise minimisation? To answer this question we consider the derivative of $w(x)$ with respect to the parameter $\beta$,

$$\frac{dw}{d\beta} = \frac{x^{k-1}[-(k-1)x^{n+1} + kx^n - (n-k+1)x + n - k]}{(x^n + \beta)^2}.$$

Let

$$f(x) = -(k-1)x^{n+1} + kx^n - (n-k+1)x + n - k,$$

and note that

$$f'(x) = -(n+1)(k-1)x^n + knx^{n-1} - (n-k+1),$$
$$f''(x) = -n(n+1)(k-1)x^{n-1} + kn(n-1)x^{n-2}.$$

We have $f(0) = n - k > 0, f(1) = 0, f'(1) = 0, f''(1) = n(n+1-2k)$. It follows that if $n + 1 \geqslant 2k$, then $dw/d\beta$ is nonnegative in the interval $0 \leqslant x \leqslant 1$. Hence $w$ is an increasing function of $\beta$ in this interval and $w$ is minimised by taking the lowest permissible value of $\beta$, namely $\beta = (n-k)/k$. For this value of $\beta$ we get

$$w(x) = \frac{nx^k}{kx^n + n - k}, \qquad 0 \leqslant x \leqslant 1, \qquad 2k \leqslant n + 1.$$

Of these $w(x)$ the largest value of $k$ gives the smallest $w(x)$. In the case $n = 3$ for example, the choices of $k$ are $k = 1, 2$, and since $2 \cdot 2 = 3 + 1$, i.e., $2k = n + 1$, we see that the function

$$w(x) = \frac{3x^2}{2x^3 + 1}$$

pointwise minimises all members of $C$ in the interval $0 \leqslant x \leqslant 1$. Consequently for cubics over cubics sigmoidicity is maximised by the rational function

$$\frac{3x^3}{2x^3 + 1},$$

or equivalently $x^3/(x^3+1)$ after rescaling. Similarly for quadratics over quadratics sigmoidicity is maximised by $x^2/(x^2+1)$.

When we come to the case $n=4$ a new phenomenon arises. We have three one-parameter families of $w(x)$ corresponding to $k=1,2,3$. For $k=1,2$ we have $2k \leqslant n+1$ and for these values of $k$ we saw above that

$$\frac{nx^k}{kx^n+n-k}$$

are minimising functions within each family and decrease pointwise as $k$ increases for $0 \leqslant x \leqslant 1$. It follows that the function

$$\frac{4x^3}{3x^4+1},$$

which belongs to the family $k=3$, actually minimises all members of both families $k=1,2$. However, within the family $k=3$, there is no pointwise minimising function. In fact the graphs of any two members of this family cross over at some point in the interval $0 \leqslant x \leqslant 1$. To see why this is true let us consider the general member of the family $k=3$, namely

$$\frac{2(1-\beta)x^3+(3\beta-1)x^2}{x^4+\beta},$$

where $1/3 \leqslant \beta \leqslant 1$. The difference between two members with parameters $\beta, \beta'$ of the family is given by

$$\frac{(\beta'-\beta)x^2[2x^5-3x^4+2x-1]}{(x^4+\beta)(x^4+\beta')},$$

which vanishes between $x=0$ and $x=1$ since

$$2x^5-3x^4+2x-1=(x-1)^2(x-.6573)(x-r_1)(x-r_2),$$

where $r_{1,2}=-.5786 \pm .6526i$. This turns out to be a general phenomenon for $2k>n+1$. Hence for $n \geqslant 4$ we have reached the end of what can be achieved by pointwise minimisation as regards the original problem of sigmoidicity. We can summarise the discussion above by announcing a partial improvement of Proposition 4.1.

4.2. PROPOSITION. *In the interval $0 \leqslant x \leqslant 1$ each member $u(x)$ of the class $C$ is pointwise minimised by at least one $w(x)$ of a $k$-family with $2k>n+1$, if $n \geqslant 4$. For $n=2,3$ the functions $2x/(x^2+1)$, $3x^2/(2x^3+1)$ are absolutely pointwise minimising in the class $C$ over the interval $0 \leqslant x \leqslant 1$.*

To complete this work we present a special argument to show that sigmoidicity for quartics over quartics is still maximised by the function $4x^4/(3x^4+1)$ even though the argument of absolute pointwise minimisation used for the cases $n=2,3$ will not work for $n=4$. We wish then to minimise the integral

$$\int_0^1 xu(x)\,dx,$$

and we know from earlier work that we may restrict attention to functions

$$u(x)=\frac{2(1-\beta)x^3+(3\beta-1)x^2}{x^4+\beta}$$

taken from the family $k=3$. Let us consider therefore the parametrised integral

$$I(\beta)=\int_0^1 \frac{2(1-\beta)x^4+(3\beta-1)x^3}{x^4+\beta}\,dx.$$

We differentiate under the integral sign to get

$$\frac{dI}{d\beta} = \int_0^1 \frac{x^3[-2x^5 + 3x^4 - 2x + 1]}{(x^4 + \beta)^2}\, dx,$$

and, as we saw earlier, the integrand changes sign at $x = t \approx .6573$. We break the integral for $dI/d\beta$ into parts and estimate the size of each,

$$\frac{dI}{d\beta} = \int_0^t \frac{x^3[-2x^5 + 3x^4 - 2x + 1]}{(x^4 + \beta)^2}\, dx + \int_t^1 \frac{x^3[-2x^5 + 3x^4 - 2x + 1]}{(x^4 + \beta)^2}\, dx.$$

Over the interval $0 \leqslant x \leqslant t$, the integrand is positive and over $t \leqslant x \leqslant 1$ it is negative. Hence

$$\begin{aligned}\frac{dI}{d\beta} &> \frac{1}{(t^4 + \beta)^2} \int_0^t x^3[-2x^5 + 3x^4 - 2x + 1]\, dx \\ &\quad + \frac{1}{(t^4 + \beta)^2} \int_t^1 x^3[-2x^5 + 3x^4 - 2x + 1]\, dx \\ &= \frac{1}{(t^4 + \beta)^2} \int_0^1 x^3[-2x^5 + 3x^4 - 2x + 1]\, dx \\ &= \frac{1}{(t^4 + \beta)^2} \left[ -\frac{2}{9} + \frac{3}{8} - \frac{2}{5} + \frac{1}{4} \right] \\ &= \frac{1}{(t^4 + \beta)^2} \cdot \frac{1}{360} \\ &> 0.\end{aligned}$$

Hence $I$ is an increasing function of $\beta$ and is a minimum therefore at the least permissible value of $\beta$ namely $\beta = \frac{1}{3}$. This gives our final result for quartics over quartics (after renormalising).

4.3. PROPOSITION. *Maximum sigmoidicity for quartics over quartics is achieved by the function*

$$\frac{x^4}{x^4 + 1}.$$

**References**


1. W. G. Bardsley, The 3 : 3 function in enzyme kinetics. Possible shapes of $v/S$ and $(1/v)/(1/S)$ plots for third degree steady-state rate equations, J. Theoret. Biol. (1977) 65, 281–316.

2. W. G. Bardsley and R. E. Childs, Sigmoid curves, non-linear double-reciprocal plots and allosterism, Biochem. J. (1975) 149, 313–328.

3. W. G. Bardsley and A. J. Wright, A new approach to the measurement of sigmoid curves with enzyme kinetic and ligand binding data, J. Mol. Biol. (1983) 165, 163–182.

4. W. G. Bardsley, R. Woolfson, and J.-P. Mazat, Relationships between the magnitude of Hill plot slopes, apparent binding constants and factorability of binding polynomials and their Hessians, J. Theoret. Biol. (1980) 85, 247–284.

5. W. G. Bardsley, R. Woolfson, and R. M. W. Wood, Some mathematical results concerning Hessians of binding polynomials and cooperativity coefficients, J. Theoret. Biol. (1980) 85, 45–51.

6. D. E. Koshland, Jr., G. Némethy, and D. Filmer, Comparison of experimental binding data and theoretical models in proteins concerning subunits, Biochemistry (1966) 5, 365–385.

7. J. Monod, J. Wyman, and J.-P. Changeux, On the nature of allosteric transitions: A plausible model, J. Mol. Biol. (1965) 12, 88–118.

8. I. Todhunter, An elementary treatise on the theory of equations, Macmillan, 1882.

9. J. M. Wardell, W. G. Bardsley, J. P. Kavanagh, and R. M. Wood, The probability of obtaining complex kinetic curves for enzyme mechanisms with cubic terms in the pseudo-steady state rate equations, J. Theoret. Biol. (1982) 95, 465–487.

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S, L**.** Martin Gardner's Sixth Book of Mathematical Diversions from Scientific American. Martin Gardner. U of Chicago Pr, 1983, x + 262 pp, $7.95 (P). [ISBN: 0-226-28250-3] Paperback reprint of the well-known 1971 W.H. Freeman collection, with the word "Games" changed to "Diversions" in the title. LAS

**General, S, L**.** Martin Gardner's New Mathematical Diversions from Scientific American. Martin Gardner. U of Chicago Pr, 1983, 253 pp, $7.95 (P). [ISBN: 0-226-28247-3] Reprint of the 1966 Simon and Schuster original edition (with some updates, e.g., on the 1976 proof of the four-color conjecture). A classic, now available for a new generation of students. LAS

**Elementary, T(13: 1).** Introductory Geometry. Mary Kay Hudspeth. Addison-Wesley, 1983, 557 pp, $18.95 (P). [ISBN: 0-201-10690-6] This typescript text covers the fundamentals of high school geometry and introduces formal logic and trigonometry; intended for developmental students with knowledge of first year algebra, but presumes no prior knowledge of geometry. JNC

**Mathematics Appreciation, S???(13).** Generalisations et Generalites. Florentin Smarandache. Edition Nouvelle, 1984, 53 pp, (P). Trivial generalizations of known results, plus a collection of the author's favorite mathematical oddities. JRG

**Precalculus, T(13-14: 2-4).** Basic Technical Mathematics. Peter K.F. Kuhfittig. Brooks/Cole, 1984, xvi + 685 pp, $24.95 [ISBN: 0-534-03074-2]; Basic Technical Mathematics with Calculus, xxiii + 1030 pp, $28.95. [ISBN: 0-534-03151-X] Designed for students in technical and engineering programs, both texts make use of technical applications and notation and include calculator procedures and problems. The first text covers standard precalculus topics including some analytic geometry. The second text includes all the material from the first text as well as first year calculus. JNC

**Precalculus, T(13: 1).** Trigonometry. Charles P. McKeague. Academic Pr, 1984, xv + 458 pp, $20. [ISBN: 0-12-484785-4] Standard trigonometry text; includes complex numbers and appendix on logarithms. Many exercises. RM

**Precalculus, T*(13: 1).** Precalculus: Functions & Graphs. Bernard Kolman, Arnold Shapiro. Academic Pr, 1984, xiv + 663 pp, $21. [ISBN: 0-12-417894-4] A careful presentation of the standard precalculus topics emphasizing functions and graphing techniques; each chapter includes one or two "feature" applications, interspersed progress checks and warnings, and concludes with lists of terms, symbols and key ideas as well as review exercises and two progress tests. JNC

**Education, T(13: 1).** Modern Elementary Mathematics, Fourth Edition. Malcolm Graham. HarBrace J, 1984, xii + 575 pp, $25.95. [ISBN: 0-15-561043-0] Primarily for pre-service teachers of K-8; a new chapter on mathematical reasoning and problem solving precedes material on sets, functions, number systems, elementary geometry (including rigid motions) and some statistics. Other revisions include additional illustrations and problems and some calculator exercises. (First Edition, TR, January 1971; Extended Review, January 1972; Second Edition, TR, October 1975; Third Edition, TR, November 1979.) JNC

**Education, T(13: 1).** Modern Mathematics: An Elementary Approach, Sixth Edition. Ruric E. Wheeler. Brooks/Cole, 1984, xv + 694 pp, $23.95. [ISBN: 0-534-02843-8] For elementary education and liberal arts majors; covers problem solving, elementary number theory, geometry, probability and statistics, calculators and computers; new features include discussion of cognitive reasoning techniques and "personal" notes to the students introducing each section. Also new is an accompanying Student Manual. (Second Edition, TR, November 1970; Third Edition, TR, October 1973; Fourth Edition, TR, December 1977; Alternate Edition, March 1982; Fifth Edition, November 1982.) JNC

**History, P*, L*.** A History of Mathematics Education in England. Geoffrey Howson. Cambridge U Pr, 1982, x + 294 pp, $54.50. [ISBN: 0-521-24206-1] The first account of mathematics education in Eng-

land, set as a series of nine biographies of representative English mathematics educators, spanning the period from 1530 to 1960. These effective and interesting profiles reveal not only individual accomplishments, but also the educational and cultural context of each person's life. A prelude and a postlude offer highlights of events preceding 1500 and following 1960; appendices contain sample examinations and syllabi. Although the biographical approach is popular in tone, copious notes provide scholars with detailed pointers to primary sources. LAS

**History, L**.** Understanding Relativity: Origin and Impact of a Scientific Revolution. Stanley Goldberg. Birkhauser Boston, 1984, xvii + 494 pp, $24.95. [ISBN: 3-7643-3150-X] Explication (for the lay reader) of the special and general theories, followed by description and discussion of early response to and eventual assimilation of the theory in Europe and the U.S. Interesting discussion of sociological implications of this major scientific advance. Technical material in appendices. Bibliographic essay. JRG

**Combinatorics, T*(15-16: 1, 2), S*, P, L*.** Applied Combinatorics. Fred S. Roberts. Prentice-Hall, 1984, xviii + 606 pp, $37.95. [ISBN: 0-13-039313-4] A versatile text which stresses algorithmic thinking and provides illustrations and applications to a wide variety of fields. Subject matter includes elementary combinatorics, graphs, generating functions, recurrences, inclusion/exclusion, Pólya theory, experimental design, coding, combinatorial optimization. Large collections of exercises. LCL

**Number Theory, S(15-18), L*.** Number Theory: An Approach Through History From Hammurapi to Legendre. André Weil. Birkhauser Boston, 1984, xxi + 375 pp, $24.95. [ISBN: 3-7643-3141-0] A thoroughly engaging treatise covering the history of number theory until 1801. The three principal chapters focus on Fermat, Euler, and Lagrange and Legendre. Appendices to chapters deal with such topics as Mordell's theorem, addition theorem for elliptic curves, quadratic reciprocity, and quadratic forms. A must for any teacher of number theory at any level. SG

**Number Theory, S(13-16), L.** Number Theory in Science and Communication: With Applications in Cryptography, Physics, Biology, Digital Information, and Computing. M.R. Schroeder. Ser. in Inform. Sci., V. 7. Springer-Verlag, 1984, xvi + 324 pp, $24.50. [ISBN: 0-387-12164-1] An informal introduction to some applications of elementary number theory, primarily in coding but with some less publicized examples, e.g., fast Fourier transforms. Some previous acquaintance with number theory useful but not essential. Good chapter-by-chapter references to more in-depth articles in the literature. DD

**Number Theory, S, P, L*.** Factorizations of $b^n+1$, $b = 2,3,5,6,7,10,11,12$ up to high powers. John Brillhart, et al. Contemporary Math., V. 22. AMS, 1983, lxviii + 178 pp, $22 (P). [ISBN: 0-8218-5021-0] A book of tables which gives known factorizations of numbers of the form $b^n+1$ for the values of b mentioned in the title. Also includes an historical account of hardware and theoretical development and a list of references. CEC

**Linear Algebra, T(13-14: 1), L.** Introductory Linear Algebra with Applications. John W. Brown, Donald R. Sherbert. Prindle, Weber & Schmidt, 1984, xi + 436 pp. [ISBN: 0-87150-700-5] An interesting departure in both content and spirit from most linear algebra books, this freshman-sophomore text grew out of a revision of the typical finite mathematics course in order to give a clearer perspective and emphasis on the linear algebra portions of it. Theory is stated but largely unproved. Emphasis is on applications, including linear programming, systems of equations, Markov chains, least squares, and generalized inverses. Exercises, answers, index. JS

**Linear Algebra, T**(13-14: 1, 2).** Elementary Linear Algebra, Second Edition. Stanley I. Grossman. Wadsworth, 1984, xv + 426 pp, $30. [ISBN: 0-534-02738-5] Changes include the rewriting of some sections to improve clarity and to promote understanding. Format changes in uses of color and print. An attractive book. Provides many options for depth and extent of coverage. Have a look. (First Edition, TR, April 1980.) JK

**Linear Algebra, T(14-16: 1), L.** Linear Algebra with Applications. Gareth Williams. Allyn & Bacon, 1984, viii + 504 pp, $30. [ISBN: 0-205-08010-3] Includes an adequate introduction to the elementary theory of linear algebra but the emphasis is largely toward applications and computational aspects. Wider than usual choice of examples, in such areas as population problems, Markov chains, coding theory, linear programming, and computer solutions. Exercises, answers, appendices. JS

**Group Theory, P.** Finite Group Algebras and their Modules. P. Landrock. London Math. Soc. Lect. Note Ser., V. 84. Cambridge U Pr, 1983, x + 274 pp, $27.50 (P). [ISBN: 0-521-27487-7] Text on modern representations and group algebras in characteristic p, concentrating on "post-Brauer" results. DD

**Group Theory, S(17-18), P.** American Mathematical Society Translations, Series 2, Volume 121: The Kourovka Notebook: Unsolved Problems in Group Theory, Seventh Augmented Edition. AMS, 1983, v + 112 pp, $33. [ISBN: 0-8218-3079-1] This Seventh Edition contains the complete list of 422 problems from the Sixth Edition (151 have been solved) together with update comments, bibliographic information, etc., and an additional chapter with 58 new problems. LCL

**Algebra, T*(15-16: 1), S*, P, L*.** Discrete Mathematics and Applied Modern Algebra. Henry B. Laufer. Prindle, Weber & Schmidt, 1984, vi + 538 pp. [ISBN: 0-87150-702-2] A course in applied modern algebra (groups, Boolean algebra, rings and fields) with motivation and applications from coding theory, fast-arithmetic, finite state machines, graph theory, formal language. Pleasing format, easy to

read text, many examples and exercises. LCL

**Algebra, T(17), S, L.** Algebra. Larry C. Grove. Pure & Appl. Math. Academic Pr, 1983, xiii + 299 pp, $32. [ISBN: 0-12-304620-3] Suitable for a first-year graduate course or for independent reading by an advanced undergraduate. The text covers the basics of groups, rings, fields, modules, and algebras. The approach is classical with no mention of category theory or homological methods. SG

**Algebra, T(18: 1), S, P.** Infinite Dimensional Lie Algebras, An Introduction. Victor G. Kac. Prog. in Math., V. 44. Birkhauser Boston, 1983, xvi + 245 pp, $18.75. [ISBN: 0-8176-3118-6] A detailed, largely self-contained and comprehensive treatment of the present state of development of the class of infinite-dimensional Lie algebras now known as Kac-Moody algebras; particular attention to affine Lie algebras and highest weight modules. Exercises, notes and commentary, bibliography, index. JS

**Algebra, T(18: 1), S, P.** Skew Fields. P.K. Draxl. London Math. Soc. Lect. Note Ser., No. 81. Cambridge U Pr, 1983, ix + 182 pp, $19.95 (P). [ISBN: 0-521-27274-2] Based on lecture notes given at Penn State in 1981, there are three parts: I. Skew Fields and Simple Rings; II. Brauer Groups; and III. K-Theory. Some exercises, bibliography. JS

**Algebra, P.** Lecture Notes in Mathematics-1050: Formally p-adic Fields. Alexander Prestel, Peter Roquette. Springer-Verlag, 1984, v + 167 pp, $9.50 (P). [ISBN: 0-387-12890-5] Formally p-adic fields are fields of characteristic zero having at least one p-valuation. Investigates all p-adic valuations on a given field, the model theory of such fields (deducing the Ax-Kochen-Ersov Theorem) and considers in particular function fields over p-adically closed fields. DD

**Algebra, P.** Commutative Semigroup Rings. Robert Gilmer. Lect. in Math. U of Chicago Pr, 1984, xii + 380 pp, $11 (P); $27. [ISBN: 0-226-29392-0; 0-226-29391-2] Studies the structure of commutative semigroup rings, e.g., polynomial rings over commutative coefficient rings, asking when such an object is a Krull domain, when two are isomorphic, etc. DD

**Algebra, T(17-18: 1), S, P.** A First Course of Homological Algebra. D.G. Northcott. Cambridge U Pr, 1980, xi + 206 pp, $13.95 (P). [ISBN: 0-521-20196-9] Paperback reprint of the 1973 hardcover original edition (TR, February 1974). LAS

**Finite Mathematics, T(13-15: 1, 2).** Applied Finite Mathematics. Gary G. Gilbert, Donald O. Koehler. McGraw-Hill, 1984, xix + 549 pp, $27.50. [ISBN: 0-07-023222-9] Emphasis on mathematical modelling. Many examples and applications adapted from research articles in applied areas. Includes discussion of errors in the modelling process. Many exercises, from routine to real applications. Topics include linear programming, graphs, probability and Markov chains, game theory, finance. RM

**Calculus, T(13: 3).** Calculus with Analytic Geometry, Second Edition. M.A. Munem, D.J. Foulis. Worth, 1984, xvi + 1048 pp, $37.95. [ISBN: 0-87901-236-6] Changes from the First Edition (TR, November 1978) include adding several new topics and some computer-generated graphs, de-emphasizing the use of tables, and using metric units in the applications. Two study guides and a computer workbook and diskette are also available. BK

**Calculus, T*(13).** Essentials of Calculus for Business, Economics, Life Sciences, Social Sciences. Louis Leithold. Harper & Row, 1984, xv + 556 pp, $30.50. [ISBN: 0-06-043954-8] Concise, clear statements of definitions and theorems (without proofs) and careful explanations of concepts are motivated and illustrated by applications from the indicated areas; includes chapters on exponential and logarithmic functions and on differential calculus of functions of several variables; an appendix covers the calculus of trigonometric functions. JNC

**Calculus, T(13: 1, 2), S.** Algebra and Fundamentals of Analysis. Ed: G.N. Yakovlev. Transl: Leonid Levant. MIR, 1981, 567 pp, $12.95. [ISBN: 0-8285-2571-4] A textbook for students at secondary technical schools and colleges in the USSR. Standard fare in elementary mathematical analysis, including chapters on computational mathematics and basics of probability. Numerous worked-out examples. Almost 700 problems, all with answers. Physically compact and inexpensive. Suitable for self-study. A good buy. JK

**Calculus, T*(13-14: 1-3).** Calculus and Analytic Geometry. Donald W. Trim. Addison-Wesley, 1983, xix + 923 pp, $37.95. [ISBN: 0-201-16270-9] Attractive textbook featuring lots of exercises (some requiring a calculator) and refreshing figures. Usual fare but the author's informal style and intuitive approach, the variation in the order of topics and the recasting of standard exercises enhance this presentation. Emphasis is on problem solving through an understanding of concepts. Answers are given to the even-numbered exercises. JK

**Calculus, T(13-14: 1-3).** Calculus, Third Edition. Stanley I. Grossman. Academic Pr, 1984, xxiii + 1336 pp, $31. [ISBN: 0-12-304371-9] Changes in examples, exercises, theorems and explanations to improve clarity. Some rearrangement of topics. Trigonometry now appears earlier in the text. Detailed biographical sketches have been added. Continues to be a solid textbook. (First Edition, June-July 1979; Second Edition, TR, November 1981.) JK

**Calculus, T*(13: 2, 3), S, L.** Calculus with Analytic Geometry, Fourth Edition. Edwin J. Purcell, Dale Varberg. Prentice-Hall, 1984, xv + 864 pp, $36.95. [ISBN: 0-13-111807-2] Brief but brisk and to the point. Emphasizes functions as main objects of study; differentiation and integration as linear operators on functions. Trigonometric functions are reviewed early. Full chapter on elementary numerical methods. Exercises unusually varied and imaginative; nearly all sets include graphi-

cal or calculator problems (Third Edition, TR, October 1978). PZ

**Calculus, T(113: 2, 3), S.** Calculus and Analytic Geometry, Fourth Edition. Douglas F. Riddle. Wadsworth, 1984, xvii + 1246 pp, $43.25. [ISBN: 0-534-01468-2] Set of topics is standard. First chapters relatively elementary, intuitive; rigor increases later. Limits, continuity treated in three widely separated chapters--first intuitively, then more rigorously in geometric setting, finally in epsilon-delta format. Transcendental functions follow fundamental theorem. General emphasis on geometry, in text and exercises. (First Edition, TR, April 1970; Second Edition, TR, October 1974; Third Edition, TR, March 1979.) PZ

**Calculus, S***(13-14).** Interface: Calculus and the Computer, Second Edition. David Smith. Saunders College, 1984, xv + 254 pp, (P). [ISBN: 0-03-070663-7] This book is a first in many ways, even though it is a Second Edition (First Edition, TR, November 1976). It is not the first well-written book about calculus, though it is among the rare few; neither is it the first book to treat the computer as a tool for enhancing calculus. But it is the first to bring the two together with so much delight and substance. It teaches BASIC (with appropriate apologies) on the sly while guiding the reader, who might profitably be a calculus teacher, to new insights into calculus and computing. It focuses on numerical approaches that develop understanding by requiring programming; it only mentions symbolic manipulation. But the ideas presented go definitely beyond those of numerical methods into history and modelling. After the very first chapters where pidgin BASIC is developed, the rest of the algorithms are presented in a language independent flow-chart style. Thus the book is suitable in conjunction with any machine as powerful as the more sophisticated programmable calculators. Good references, indexes, reference charts and the like. JAS

**Calculus, T(13-14: 1).** Calculus for Business. Richard S. Paul, Ernest F. Haeussler, Jr. Reston, 1984, vii + 374 pp, $23.95. [ISBN: 0-8359-0635-3] Strictly for business. "Technical proofs, conditions, etc., are sufficiently described but are not overdone;" omissions include mean value theorem, volumes of revolution, proof of chain rule. Self-contained with respect to economic terms. Thirty-page reference appendix on algebra, with exercises. Attention is drawn to common mistakes in "pitfall" paragraphs. RB

**Calculus, T(13).** Calculus: Basic Concepts and Applications. R.A. Rosenbaum, G. Philip Johnson. Cambridge U Pr, 1984, xvi + 422 pp, $24.95. [ISBN: 0-521-25012-9] An intuitive calculus text concentrates on a single variable, and achieves brevity by using short problem lists, plausibility instead of proof, and fast moving prose. Attractive possibility for a short course. AWR

**Real Analysis, S(16-17), P, L.** Asymptotic Analysis. J.D. Murray. Appl. Math. Sci., V. 48. Springer-Verlag, 1984, vii + 164 pp, $22. [ISBN: 0-387-90937-0] Written mainly for the person whose "primary aim is to get answers to practical problems," this is a readable, nonrigorous approach to asymptotic solutions for various methods including steepest descent, stationary phase, integral transitions, differential equations, and singular perturbations. Includes exercises, bibliography, index. (1974 Oxford University Press original edition, TR, November 1975.) JS

**Complex Analysis, T(16-17: 1, 2), L.** The Theory of Analytic Functions: A Brief Course. A.I. Markushevich. Transl: Eugene Yankovsky. MIR, 1983, 423 pp, $10.95. Introduction to standard topics of complex analysis at advanced undergraduate or beginning graduate level. Emphasizes geometric, physical, engineering points of view in interpreting results. Relatively complete: Nevanlinna theory, quasiconformal mappings, and several-variable theory are introduced. Many diagrams, but almost no exercises. PZ

**Differential Equations, P.** Hyperbolic Partial Differential Equations: Populations, Reactors, Tides and Waves: Theory and Applications. Ed: Matthew Witten. Intern. Ser. in Modern Appl. Math. & Computer Sci., V. 6. Pergamon Pr, 1983, viii + 244 pp, $37.50. [ISBN: 0-08-030254-8] A retitled special issue of Vol. 9, No. 3 of the journal Computers and Mathematics with Applications, containing papers devoted primarily to the McKendrick/Von Foerster hyperbolic partial differential equations model of population growth. LAS

**Differential Equations, T(15-16: 1), S*, L.** Ordinary Differential Equations in $R^n$: Problems and Methods. L.C. Piccinini, G. Stampacchia, G. Vidossich. Transl: A. LoBello. Appl. Math. Sci., V. 39. Springer-Verlag, 1984, xii + 385 pp, $32 (P). [ISBN: 0-387-90723-8] Excellent readable, contemporary account of topics usually covered in differential equations in a course other than the first. Based on the notes of G. Stampacchia. Exercises extend and complement the text. Not appropriate for students in the U.S. who have completed only the usual sequence of calculus courses. A very attractive treatment of the material. JK

**Differential Equations, S(18), P.** Almost Periodic Functions and Differential Equations. B.M. Levitan, V.V. Zhikov. Transl: L.W. Longdon. Cambridge U Pr, 1982, xi + 211 pp, $34.50. [ISBN: 0-521-24407-2] Develops general properties of almost periodic functions and applies this theory to linear and non-linear operator differential equations. Clearly written translation from the Russian. Historical references after each chapter. No problems. BH

**Differential Equations, T*(17: 3), S*, P*, L.** Elliptic Partial Differential Equations of Second Order, Second Edition. David Gilbarg, Neil S. Trudinger. Grund. der math. Wissenschaften, B. 224. Springer-Verlag, 1983, xiii + 513 pp, $45.50. [ISBN: 0-387-13025-X] Covers the theory of the linear and non-linear elliptic Dirichlet problem, from the classical and modern points of view. Pays particular attention to the equations of mean curvature, minimal surface, and Monge-Ampère. Provides the necessary background in Sobolev spaces, fixed point theory, and norm estimates. Mentions appli-

cations, e.g., stochastic control theory. Offers historical notes and problems. May be accessible to advanced undergraduates. (First Edition, TR, August-September 1978.) YN

**Numerical Analysis, S(13).** Applied Numerical Methods for the Microcomputer. Terry E. Shoup. Prentice-Hall, 1984, x + 262 pp, $25.95 [ISBN: 0-13-041418-2]; Numerical Methods for the Personal Computer, 1983, xii + 238 pp, $18.95. [ISBN: 0-13-627208-8] Standard numerical methods are presented for the solution of engineering-oriented problems on a microcomputer. Programs for many of the methods are listed in a particular version of BASIC. Some general discussion of microcomputers and choice of methods. RWN

**Numerical Analysis, T(14-15).** Introduction to Numerical Computation. J. Thomas King. McGraw-Hill, 1984, xi + 355 pp, $31.95. [ISBN: 0-07-034639-9] A textbook for an undergraduate course in numerical analysis. Covers standard topics: the solution of nonlinear equations, polynomial interpolation, elimination methods for systems of linear equations, numerical integration and differentiation, and discrete variable methods for initial value problems. Example programs are written in WATFIV. AO

**Numerical Analysis, T(14-15: 1, 2), L.** Applied Numerical Analysis, Third Edition. Curtis F. Gerald, Patrick O. Wheatley. Addison-Wesley, 1984, x + 621 pp, $29.95. [ISBN: 0-201-11577-8] Major changes from earlier editions include the rewriting of all computer programs (programs are now in Fortran 77), many new exercises, as well as new material on trigonometric approximation, the Fast Fourier Transform, and adaptive integration. AO

**Numerical Analysis, T*(17-18: 1), P.** Matrix Computations. Gene H. Golub, Charles F. van Loan. Ser. in the Math. Sci., V. 3. Johns Hopkins U Pr, 1983, xvi + 476 pp, $49.50; $24.95 (P). [ISBN: 0-8018-3010-9; 0-8018-3011-7] A clear, relatively comprehensive text on numerical linear algebra, including linear systems, least squares and eigenvalue problems. Presents approaches that have been of recent value for large matrices, error and sensitivity analyses, and special methods. RWN

**Numerical Analysis, T(15-16: 1, 2).** Introduction to Numerical Computations, Second Edition. James S. Vandergraft. Computer Sci. & Appl. Math. Academic Pr, 1983, xiv + 372 pp, $29.50. [ISBN: 0-12-711356-8] Two topics have been added in this edition: Newton formulas for interpolation, and adaptive methods for integration. (First Edition, TR, August-September 1979.) AO

**Functional Analysis, T(18: 1), S, P.** Locally Convex Spaces and Operator Ideals. Heinz Junek. Teubner-Texte zur Math., B. 56. BG Teubner, 1983, 180 pp, 17M (P). A systematic approach to the study of locally convex vector spaces by way of looking at ideals of operators. Special attention is paid to (F)-, (DF)-, and nuclear spaces. Assumes background in Banach and Hilbert Space. References, index. JS

**Functional Analysis, P.** Approximation of Hilbert Space Operators, Volume II. Constantin Apostol, et al. Research Notes in Math., No. 102. Pitman, 1984, 524 pp, $29.95 (P). [ISBN: 0-273-08641-3] Chapters 9-16, plus appendices, continuing the work begun by Domingo Herrero in Volume I. Emphasizes norm-approximation problems of invariant subsets of operator algebras. LAS

**Functional Analysis, T**(18: 1), S**, P**.** Sequences and Series in Banach Spaces. Joseph Diestel. Grad. Texts in Math., No. 92. Springer-Verlag, 1984, xiii + 261 pp, $38. [ISBN: 0-387-90859-5] An engaging introduction, in 14 brief chapters, to results and methods of structure theory of Banach spaces. Focuses on recent developments but gives consistent attention to history of the topic. Style is informal, readable, contagiously enthusiastic, and generous to the reader--helpful comments and paraphrases of important results abound. Exercises and notes follow each chapter. PZ

**Functional Analysis, P.** Lecture Notes in Mathematics-1012: Spectral Theory of Banach Space Operators. Shmuel Kantorovitz. Springer-Verlag, 1983, 179 pp, $9.50 (P). [ISBN: 0-387-12673-2]

**Functional Analysis, P.** Lecture Notes in Mathematics-1049: Cônes autopolaires et algèbres de Jordan. Bruno Iochum. Springer-Verlag, 1984, vi + 247 pp, $13 (P). [ISBN: 0-387-12901-4] On the underlying order structure of a Jordan algebra. Of interest to quantum theorists. JD-B

**Functional Analysis, P.** Infinitely Divisible and Stable Measures on Banach Spaces. Werner Linde. Teubner-Texte zur Math., B. 58. BG Teubner, 1983, 201 pp, 19M (P). Spectral representations, tail behavior, and other properties of stable and infinitely divisible measures on Banach spaces. An infinitely divisible (Radon) measure admits n-th roots for all n--such a measure is also stable if each root is a translation and dilation of the original measure. The subject is a branch of probability theory on Banach spaces. PZ

**Analysis, S(16-17), L.** Functional Equations: History, Applications and Theory. Ed: J. Aczél. Math. & Its Applic. D Reidel, 1984, ix + 244 pp, $47.50. [ISBN: 90-277-1706-0] Part of a series dedicated to showing relationships between apparently disparate areas of mathematics, this book collects papers and panel contributions from the functional equations section of a conference, Mathematics at the Service of Man. Its value is greatly enhanced by attention to historical development, though the pleasure of such reading is diminished by the strained, sometimes fractured English prose of writers more at home with other languages. AWR

**Analysis, S*(18), P.** Geometrical Methods of Nonlinear Analysis. M.A. Krasnosel'skiĭ, P.P. Zabreĭko. Grund. der math. Wissenschaften, B. 263. Springer-Verlag, 1984, xix + 409 pp, $48. [ISBN: 0-387-12945-6] Written for "mathematicians and mechanical engineers" with a "basic knowledge of functional analysis and topology." Describes theory of rotation of vector fields and applies this theory to

the study of solution sets of operator equations, giving many mathematical and physical examples. Text is well motivated, clearly written, and provides good overview of material as it is developed. Unfortunately, no problems. BH

**Analysis, P.** Inégalités isopérimétriques et applications en physique. Jacqueline Mossino. Hermann, 1984, 182 pp, 130 F (P). [ISBN: 2-7056-5963-3]

**Algebraic Geometry, P.** Intersection Theory. William Fulton. Ergebnisse der Math. und ihrer Grenzgebiete, 3. Folge, Band 2. Springer-Verlag, 1984, xi + 470 pp, $39. [ISBN: 0-387-12176-5] A comprehensive modern treatment of the subject. Serves well as a text for graduate students in algebraic geometry or as a basic reference for practicing algebraic geometers. SG

**Differential Geometry, P.** Séminaire Sud-Rhodanien de Géométrie. P. Dazord, N. Desolneux-Moulis. Hermann, 1984. Feuilletages et quantification géométrique, 139 pp, 130 F (P). [ISBN: 2-7056-5972-2]; Géométrie symplectique et de contact, 113 pp, 110 F (P). [ISBN: 2-7056-5971-4] The first two of three volumes containing papers given at a conference on symplectic geometry in Lyons in June 1983. JD-B

**Differential Geometry, P.** Affine Differential Geometry. Su Buchin. Gordon & Breach, 1983, iv + 248 pp, $63.75. [ISBN: 0-677-31060-9] From Shanghai: the study of geometric objects invariant under affine transformations, after Blaschke and the author's own work of the 1920's and '30's. Review of Blaschke's Vorlesungen, affine surface theory, relations between affine and projective geometry. Presumes elementary differential geometry (curves and surfaces). Exercises. RB

**Differential Geometry, P.** Introduction to the Geometry of Foliations, Part B: Foliations of Codimension One. Gilbert Hector, Ulrich Hirsch. Aspects of Math., V. 1. Friedr. Vieweg & Sohn (Distr: Heyden & Son), 1983, x + 298 pp, $22 (P). [ISBN: 3-528-08568-1] An extension of Part A (published 1981) to extend the geometry of foliations of surfaces to general manifolds of codimension one. JAS

**Differential Geometry, P.** Category Theoretic Methods in Geometry. Ed: A. Kock. Pub. Ser., No. 35. Aarhus U, 1983, vi + 290 pp, (P). Proceedings of the workshop which took place in Aarhus, June 1-15, 1983. JAS

**Differential Geometry, T(17: 1, 2), S, P*.** Modern Geometry--Methods and Applications, Part I: The Geometry of Surfaces, Transformation Groups, and Fields. B.A. Dubrovin, A.T. Fomenko, S.P. Novikov. Transl: Robert G. Burns. Grad. Texts in Math., V. 93. Springer-Verlag, 1984, xv + 464 pp, $48. [ISBN: 0-387-90872-2] First in a projected three-volume text for a modern course in geometry appropriate for physicists. Excellent preface contains authors' thoughts on what should be compulsory subjects in a university-level mathematical education. Topics selected with a view toward their utility include basic notions of surface theory, the algebra and the differential calculus of tensors, and the calculus of variations. Done in "as simple and concrete a language possible." Smoothly translated. Some examples and exercises. Brief bibliography. JK

**Differential Geometry, P.** Lecture Notes in Mathematics-1045: Differential Geometry. Ed: A.M. Naveira. Springer-Verlag, 1984, viii + 194 pp, $10.50 (P). [ISBN: 0-387-12882-4] Proceedings of a symposium held at Peñsicola, Spain in 1982. JG

**Differential Geometry, P.** Proceedings of the 1980 Beijing Symposium on Differential Geometry and Differential Equations. Ed: S.S. Chern, Wu Wen-tsün. Gordon & Breach, 1982, $325 set [ISBN: 0-677-31000-5]; Volume 1, ix + 625 pp; Volume 2, v + 474 pp; Volume 3, v + 640 pp. Papers from a month-long international symposium, the first of its kind ever held in China. Volumes 1 and 2 contain the invited "comprehensive reports," including papers by M.F. Atiyah, E. Bombieri, Gu Chao-hao, L.K. Hua, P.D. Lax, H. Wu, and Wu Wen-tsün. Volume 3 contains about 60 contributed papers, mostly by Chinese mathematicians. LAS

**Geometry, S(13), L.** Experiments in Four Dimensions. David L. Heiserman. TAB Books, 1983, vii + 501 pp, $17.50 (P). [ISBN: 0-8306-1541-5] A detailed "how-to" guide for generating the specifications (i.e., vertex coordinates, angles, edge lengths, etc.) of four-dimensional objects and three-dimensional views of these objects; painstakingly leads the reader with knowledge of precalculus through coordinate systems, algebraic representations of transformations and object specifications in one, two, three, and four dimensional space; exercises. JNC

**Algebraic Topology, T(17-18: 1), S, P.** Techniques of Geometric Topology. Roger A. Fenn. London Math. Soc. Lect. Note Ser., No. 57. Cambridge U Pr, 1983, viii + 280 pp, $24.95 (P). [ISBN: 0-521-28472-4] Illustrates basic techniques of algebraic topology as applied to geometric aspects of low dimensional topology (dimension less than or equal to about 4). Reviews of homology, homotopy; covering spaces including their homology; knot theory; Massey products; exercises. A suitable sequel to a semester course in algebraic topology or for independent reading, emphasizing both foundational and current geometric aspects. RB

**Algebraic Topology, T(18: 1), P.** Spectra and the Steenrod Algebra: Modules Over the Steenrod Algebra and the Stable Homotopy Category. H.R. Margolis. Math. Lib., V. 29. Elsevier Sci, 1983, xix + 489 pp, $65. [ISBN: 0-444-86516-0] "[T]he book constitutes both a new presentation of results in the literature--in effect, an introduction to stable homotopy theory--and a presentation of original results [by the author]." Focuses on global structure of topological categories, as opposed to specific problems. RB

**Algebraic Topology, T(15-17: 1), P, L.** The Mathematical Theory of Knots and Braids: An Introduction. Siegfried Moran. Math. Stud., V. 82. Elsevier Sci, 1983, xii + 295 pp, $42.50 (P). [ISBN: 0-444-86714-7] Knot theory developed via the theory of braids, after Artin, Birman. Largely self-contained, including discussions of group presentations, fundamental group, Seifert-van Kampen theorem, etc.; some knowledge of algebra and topology is presumed. Stress on algebraic aspects of the subject in general and Alexander matrices of knots and links in particular. RB

**Optimization, T, S, P.** Multicriterion Optimization in Engineering with FORTRAN Programs. Andrezj Osyczka. Transl: B.J. Davies. Ser. in Engin. Sci. Halsted Pr, 1984, viii + 178 pp, $51.95. [ISBN: 0-470-20019-7] Written with the object of making theoretical results of multicriterion optimization accessible to practicing engineers. Attempt is made to keep mathematical demands on the reader to a minimum, and concrete examples are used throughout. No problems or exercises. AWR

**Probability, T(17), P.** Nonlinear Filtering and Smoothing: An Introduction to Martingales, Stochastic Integrals and Estimation. Venkatarama Krishnan. Wiley, 1984, xiii + 314 pp, $34.95. [ISBN: 0-471-89840-6] Written for "engineering graduate students with a basic knowledge of probability theory." Refers many proofs to other texts. Stresses physical understanding and applications. BH

**Statistics, S(13-16), L.** The Business of Risk. Peter G. Moore. Cambridge U Pr, 1983, x + 244 pp, $39.50; $9.95 (P). [ISBN: 0-521-24174-X; 0-521-28497-X] A general discussion of methods to assess and handle risk, with specific examples from the fields of commerce, finance, industry, medicine, health, and public policy. Assumes only a rudimentary knowledge of mathematics and statistics. LCL

**Statistics, T(16-18: 1), S, P, L.** Pattern Recognition Approach to Data Interpretation. Diane D. Wolff, Michael L. Parsons. Plenum Pr, 1983, xiii + 223 pp, $29.50. [ISBN: 0-306-41302-7] Chapter one discusses choices of data and organization, choices of tools and their applications, philosophies of interpreting results. Chapter two provides some mathematical foundations for the techniques included. Chapter three presents three major packages: SPSS, BMDP, ARTHUR. The last chapter contains natural science applications. References. Appendices. Subject index. Computer program index. RJA

**Statistics, S(17), P.** Linear Statistical Analysis of Discrete Data. Mikel Aickin. Wiley, 1983, xvi + 358 pp, $36.95. [ISBN: 0-471-09774-8] This book integrates application, computation, and theory required for statistical analysis of discrete data. Material includes log-linear models for cross-classifications, logistic regression, inference for nested models, and covariate inference. Designed to be accessible, with instruction, to those with a course in statistics beyond the elementary level, calculus, and linear algebra. MT

**Statistics, T(17: 1).** Elementary Multivariate Analysis for the Behavioral Sciences: Applications of Basic Structure. Clifford E. Lunneborg, Robert D. Abbott. Elsevier Sci, 1983, xvi + 522 pp, $39. [ISBN: 0-444-00753-9] Chapters 2-6 review matrix concepts such as matrix algebra, rank, orthogonality, and differentiation. Chapters 7-9 discuss both theory and applications of multiple regression. Chapters 10-12 are the heart of the text; they introduce multivariate distributions and the multivariate general linear model, including T-squared, Wilks' Lambda, and discriminant analysis. The final three chapters discuss models for categorical data and the basic structure and analysis of interdependence. MT

**Statistics, S(13).** Statistical Illusions: Problems. Schuyler W. Huck, Howard M. Sandler. Harper & Row, 1984, xiii + 175 pp, $10.95 (P) [ISBN: 0-06-042973-9]; Solutions, ix + 181 pp, $10.95 (P). [ISBN: 0-06-042969-0] The "problems" book contains 100 vignettes, each culminating in a leading question with an expected, "commonsense," (but incorrect) answer. The companion "solutions" volume shows why the immediate, intuitive answer is incorrect, and explains the correct solution. Can be used as a supplement in a beginning statistics course; questions span such areas as descriptive statistics, probability, estimation, hypothesis testing, correlation, nonparametrics. LCL

**Statistics, P*.** Experimental Design, Statistical Models, and Genetic Statistics: Essays in Honor of Oscar Kempthorne. Ed: Klaus Hinkelmann. Statistics, V. 50. Dekker, 1984, x + 409 pp, $55. [ISBN: 0-8247-7151-6] Collection of 21 articles written by 32 former students, colleagues and friends of Kempthorne, who did pioneer work in the areas of experimental design and genetic statistics (2 on his contributions to Iowa State University, 9 on the design and analysis of experiments, 6 on linear and nonlinear models, and 4 on statistical and population genetics). RSK

**Statistics, T(14-17), S*.** Analysis of Unbalanced Data: A Pre-Program Introduction. Ching Chun Li. Cambridge U Pr, 1982, x + 145 pp, $19.95. [ISBN: 0-521-24749-7] Can be used as a text for a short course or module to introduce the analysis of two-way classification data with unequal numbers of observations in the cells. Detailed explanations are clearly given and illustrated using only elementary algebra. An appendix gives the relation between the methods used in the text and the more sophisticated generalized inverse approach. RSK

**Statistics, S(18), L.** A Handbook of Numerical and Statistical Techniques with Examples Mainly from the Life Sciences. J.H. Pollard. Cambridge U Pr, 1981, xvi + 349 pp, $18.95 (P). [ISBN: 0-521-29750-8] Divided into three parts: Part I deals with numerical techniques, Part II deals with statistical techniques, and Part III combines numerical and statistical techniques in discussing least squares regression. This is not a text but rather a statistical handbook designed as a reference for scientists. MT

**Statistics, T(17: 2).** Intermediate Statistical Methods and Applications: A Computer Package Approach. Mark L. Berenson, David M. Levine, Matthew Goldstein. Prentice-Hall, 1983, xviii + 579 pp, $29.95. [ISBN: 0-13-470781-8] This text includes discussion of analysis of variance, regression, principal components, discriminant analysis, and analysis of cross-classified categorical data. After presentation of each statistical procedure, the book describes the commands necessary to conduct the analysis using each of three statistical computing packages: SPSS, SAS, and BMDP. MT

**Statistics, T?(14: 1), S*.** Dice, Data and Decisions: Introductory Statistics. K.W. Kemp. Ser. in Math. & Its Applic. Halsted Pr, 1984, 314 pp, $64.95. [ISBN: 0-470-27496-4] This book is an introduction to statistics for students who have had differential and integral calculus. Concepts of probability, expectation, and random variables are introduced in the first chapters, but in an order that may prove difficult to follow; one should examine this book carefully before adopting it as a text. However, there are many illustrative examples which make the book a good student reference. MT

**Statistics, T(14-17: 1, 2), P.** Biostatistical Analysis, Second Edition. Jerrold H. Zar. Prentice-Hall, 1984, xiv + 718 pp, $34.95. [ISBN: 0-13-077925-3] Presents a broad collection of data analysis techniques used by biological researchers, assuming only a background of elementary algebra. Advanced topics include parametric and nonparametric analysis of variance techniques, including multiple comparison procedures, multiple regression and correlation, and circular distributions. Contains 190 pages of tables and graphs, but no discussion of elementary probability. RSK

**Statistics, T*(15: 1, 2).** Introduction to Mathematical Statistics, Fifth Edition. Paul G. Hoel. Wiley, 1984, x + 435 pp, $28.95. [ISBN: 0-471-89045-6] Principal changes from the Fourth Edition (TR, November 1971; Extended Review, June 1974) include earlier discussion of small sample methods, a rewriting of the chapter on sampling theory, some additional material on computer programs, and many new problems, often involving real data. The problem sets now also include a final collection of mixed problems in addition to the problems grouped by section. RSK

**Computer Literacy, S, P, L.** Information, Technology and Civilization. Hiroshi Inose, John R. Pierce. WH Freeman, 1984, xx + 263 pp, $9.95 (P); $19.95. [ISBN: 0-7167-1515-5; 0-7167-1514-7] A report commissioned by the Club of Rome (ten years after their highly publicized "Limits to Growth") to examine the impact on society of information technology. The authors see information serving a new world of "communities of interest," held together by common stakes in information, data, and communication. LAS

**Computer Literacy, L.** Computer Literacy: The Basic Concepts and Language. John V. Lombardi. Indiana U Pr, 1983, x + 114 pp, $5.95 (P). [ISBN: 0-253-21075-5] A brief introduction to microcomputers and related technology for those without previous experience. Includes material on word processing, spreadsheets, communications, and programming languages. AO

**Computer Programming, T(13-14: 1).** Programming Assembler Language, Second Edition. Peter Abel. Reston, 1984, x + 581 pp. [ISBN: 0-8359-5661-X] An introduction to assembly language programming on the IBM 370 series. Includes general descriptions of the system, computer arithmetic, data handling, storage access and macros. (First Edition, TR, June-July 1980.) RWN

**Computer Programming, T(13-14: 1, 2).** PASCAL, Programming and Problem Solving. Sanford Leestma, Larry Nyhoff. Macmillan, 1984, xii + 585 pp, $19.95 (P). [ISBN: 0-02-369460-2] Intended for a two-semester introduction to Pascal programming and computer problem solving. By year's end, final chapter discusses RPN, introduction to trees, quicksort, merging: advanced programming topics of program specification, external documentation, software maintenance are excluded. Top-down problem solving demonstrated in only two examples. RB

**Computer Programming, S(13-18).** The Complete Forth. A.F.T. Winfield. Wiley, 1983, xi + 131 pp, $15.95 (P). [ISBN: 0-471-88235-6] The FORTH-79 standard dialect is used. Text directed at a microcomputer environment. First five chapters present basic language features: stack, words, actions, conditionals, loops. Next follows editing and disk operations for FORTH. Second half of text discusses specialized techniques of FORTH programming and two large complete programs. Chapter summaries and exercises. Bibliography. Answers to problems. Glossary of terminology. Index. FORTH reference card. RJA

**Computer Programming, T(13-15: 1), P.** Learning and Applying APL. B. Legrand. Transl: Julian Glyn Matthews. Wiley, 1984, xiii + 400 pp, $24.95. [ISBN: 0-471-90243-8] An introduction to the programming language APL, developed by Ken Iverson of Harvard University in 1965. APL is an operator-oriented language that is particularly well suited for solving mathematical problems relating to arrays and matrices. The text is a thorough introduction to the language and would be appropriate for either the beginner or the experienced professional wishing to learn another language. MS

**Computer Programming, T**(13-18: 1), S, L.** Paradigms and Programming with Pascal. Derick Wood. Computer Sci Pr, 1984, xiii + 425 pp, $25.95. [ISBN: 0-914894-45-5] Text is divided into four parts. First comes a quick review of Smaller Pascal, a subset of Pascal used in the text. Part two presents programming methodology issues in the context of four important problems that are examined fully. Also, there is a chapter on good I/O interfaces and one on the general issue of efficiency in high-level programming. Part three introduces a number of important paradigms (e.g., recursion, divide and conquer, various searches, backtracking, etc.) along with appropriate illustrative examples of problems and solutions. Part four very briefly introduces more extremely interesting paradigms along with project suggestions. Chapter references and exercises. Index. RJA

**Computer Programming, T*(13-18: 1).** VAX PASCAL. J.N.P. Hume, R.C. Holt. Reston, 1984, xiv + 395 pp. [ISBN: 0-8359-8247-5] Interweaves elements of computers, computer programming, the VAX programming environment, Pascal, and computer science. An introductory programming text that addresses some issues not usually found at this level; e.g., sorting and searching methods and their efficiency, insertions into and deletions from trees, other high-level programming languages as well as assembly and machine languages, compilers. Of special note are the chapters on scientific calculations and numerical methods. Chapter exercises. Several appendices. Index. RJA

**Computer Programming, T(13).** An Introduction to Programming and Pascal. William J. Collins. Macmillan, 1984, xi + 388 pp, $19.95 (P). [ISBN: 0-02-323780-5] Emphasis on programming--algorithms, pseudo-code, stepwise refinement (explained with "solution trees"). Many Pascal constructs introduced early, loops and arrays before procedures. Includes detailed case studies, but skimpy on exercises. RM

**Computer Programming, T(14-16: 1), S, P.** An Introduction to APL. S. Pommier. Transl: Bronwen A. Rees. Computer Sci. Texts, No. 17. Cambridge U Pr, 1983, vii + 136 pp, $29.95; $11.95 (P). [ISBN: 0-521-24977-5; 0-521-27109-6] APL is a function-oriented language designed by Ken Iverson. In APL, the basic data type is the array, and the language contains dozens of powerful operators for direct manipulation of these array structures. This text is a description of the APL language. It includes numerous examples of APL programs and a discussion of the types of problems for which APL is especially well suited. MS

**Computer Programming, T(15-16: 1), P, L.** Micro-PROLOG: Programming in Logic. K.L. Clark, F.G. McCabe. Prentice-Hall, 1984, xi + 401 pp, $16.95 (P). [ISBN: 0-13-581264-X] Tutorial on logic programming with the microcomputer implementation of PROLOG (available under UNIX and on Z-80 and 8088/86 based machines). Nice, straightforward treatment, with chapters on applications to expert systems, the critical path method, two person games. RM

**Computer Programming, S.** Apple BASIC Made Easy. David A. and Marianne L. Gardner. Prentice-Hall, 1984, ix + 246 pp, $19.95. [ISBN: 0-13-038928-5] A readable introduction to Apple BASIC. Includes the elementary commands, conditional statements, loops, arrays, and graphics. Some programming problems. Lots of cartoons. CEC

**Computer Programming, T(13: 1).** Introduction to Computer Science using the Turing Programming Language. R.C. Holt, J.N.P. Hume. Reston, 1984, xii + 404 pp. [ISBN: 0-8359-3168-4] Turing is a "Pascal-like" programming language that has been designed for use in teaching programming concepts to beginning programmers. It has simplified syntax, more user-friendly error messages, and simplified data structures. According to the author, it has been successfully used in a large number of introductory computer programming courses in place of either BASIC, FORTRAN, or Pascal. MS

**Computer Programming, T(13-14: 1).** Problem Solving with Structured FORTRAN 77. D.M. Etter. Benjamin/Cummings, 1984, xxii + 416 pp, $21.95 (P). [ISBN: 0-8053-2522-0] Structured problem solving with heavy emphasis on stepwise refinement, pseudocode, structured control statements, format free I/O. Many applications, chapters on file handling, character strings. RM

**Computer Programming, T(13-14: 1).** VAX FORTRAN. Charlotte Middlebrooks. Reston, 1984, xii + 516 pp, $15.95 (P). [ISBN: 0-8359-8243-2] An introduction to FORTRAN 77 intended for beginning students with no prior experience in the area of programming. The authors have chosen to make the text very machine specific. The text is geared around the DEC VAX-11 computer system. The text includes a discussion of the VAX-11 command language, editor, and filing system. Users of other computers would find this text very inappropriate. MS

**Computer Programming, P.** Programming the IBM Personal Computer: Fundamentals of BASIC. Neill Graham. Holt, Rinehart & Winston, 1984, xiii + 285 pp, (P). [ISBN: 0-03-059561-4] Introduction to BASIC on the IBM PC. Emphasis on simpler aspects, suitable for self study. RM

**Computer Programming, S(16).** INMOS Limited: OCCAM, Programming Manual. Ed: C.A.R. Hoare. Prentice-Hall, 1984, 85 pp, $16.95 (P). [ISBN: 0-13-629296-8] OCCAM is a programming language designed to support concurrent applications. It is especially useful in designing systems in which many microprocessors are networked together to solve a problem. This text is a programmer's reference manual to the OCCAM language. It assumes the reader is already familiar with the concepts of block-structured programming languages and operating systems. MS

**Computer Programming, T(14-15: 1).** Assembly Language Made Easy for the TRS-80. Chao Chien. Holt, Rinehart & Winston, 1984, xii + 228 pp, $18.45 (P). [ISBN: 0-03-070441-3] A tutorial on the TRS-80 Models I, II, III computer systems and the assembly language of the TRS-80, called EDTASM. The text assumes a familiarity with the general concepts of high level languages, and frequently uses BASIC to explain concepts or make comparisons with assembly language. MS

**Computer Programming, T(13-14).** Introduction to PASCAL and Structured Design. Nell Dale, David Orshalick. DC Heath, 1983, xvii + 586 pp, $19.95 (P). [ISBN: 0-669-06962-0] An introductory text in computer programming and the block structured language Pascal that makes no assumptions about previous experiences in either programming or mathematics. It covers the entire Pascal language and also includes discussions about such related topics as algorithms, top down design, debugging, and testing methods. MS

**Data Structures, T(15-17: 1), L.** Sequential Program Structures. Jim Welsh, John Elder, David Bustard. Ser. in Comp. Sci. Prentice-Hall, 1984, xii + 385 pp, $26.95. [ISBN: 0-13-806837-2] Focuses on the structures, both of programs and data, that arise in sequential (i.e., not concurrent) computer programs. Emphasizes the use of stepwise refinement, modular programming, abstract data types, and appropriate data structures. Stacks, queues, lists, trees, and graphs are among the data structures discussed. Example programs are presented in Pascal Plus. AO

**Software Systems, T(14-15: 1), L.** An Introduction to System Programming--Based on the PDP11. Derrick Morris. Springer-Verlag, 1983, 187 pp, $16.80 (P). [ISBN: 0-387-91230-4] Designed for an introductory course in systems programming. Covers what is necessary to make a "bare" machine usable at the assembly language level. AO

**Software Systems, L.** The Computer Graphics Glossary. Stuart W. Hubbard. Oryx Pr, 1983, vi + 95 pp, $24.50. [ISBN: 0-89774-072-6] Brief one or two sentence definitions and explanations of between 500 and 1000 terms. Mostly names of equipment and high-level concepts that would be found at the applications program level in computer graphics. JAS

**Software Systems, T*(13-18: 1), S*, P, L.** Smalltalk-80: The Language and its Implementation. Adele Goldberg, David Robson. Addison-Wesley, 1983, xx + 715 pp, $34.95. [ISBN: 0-201-11371-6] Smalltalk derives many important ideas from the simulation language Simula. The language is interpreted, interactive and object-oriented, e.g., a user can draw lines on a screen using a screen pen. Text is divided into four parts: (1) concepts and syntax of the language; (2) specification of the system's functionality; (3) modelling discrete, event-driven simulations; (4) specification of the language's virtual machine. Indexes. RJA

**Software Systems, P.** Computer Graphics: Theory and Applications. Ed: Tosiyasu L. Kunii. Springer-Verlag, 1983, x + 530 pp, $45. [ISBN: 0-387-70001-3] Proceedings of InterGraphics '83 held in Tokyo, April 11-14, 1983. These papers present the state-of-the-art in computer graphics technology and its applications. AO

**Software Systems, P.** Working with RT-11. David Beaumont, Anne Summerfield, Julie Wright. DEC, 1983, x + 206 pp, $19 (P). [ISBN: 0-932376-31-2] An introduction to the use of a single-user realtime operating system for DEC computers. First volume of four planned. RWN

**Computer Science, T(14-17: 1).** MICOS: A Microprogrammable Computer Simulator. Lubomir Bic. Computer Sci Pr, 1984, xi + 111 pp, $15.95 (P). [ISBN: 0-914894-76-5] MICOS was developed at the University of California at Irvine and is written in UCSD Pascal. It creates an environment where both the assembly language and the underlying microcode are accessible to the user. Logically the MICOS "machine" as delivered is moderately sophisticated (level of common microcomputers) and is of course microprogrammable to the users own specifications. This book describes the system and its instructions. It appears that a diskette, presumably with an applications-level manual, is available for $15 from the publisher. JAS

**Computer Science, T(16-17: 1), P.** Parallel Programming in ANSI Standard Ada. George W. Cherry. Reston, 1984, viii + 213 pp. [ISBN: 0-8359-5434-X] This text addresses the issue of parallel programming; its notation to represent these programs is ANSI standard Ada. Some of the application areas addressed are parallel sort, root finding, search, and text copying algorithms. The emphasis is on how these parallel algorithms can be effectively and efficiently implemented in Ada using the Ada rendezvous mechanism. MS

**Computer Science, S(16-17), P, L.** Fortran Optimization. Michael Metcalf. APIC Stud. in Data Processing, V. 17. Academic Pr, 1982, xii + 242 pp, $24. [ISBN: 0-12-492480-8] Presents techniques that can be used by compiler writers to generate efficient machine code from FORTRAN programs. Also discusses strategies a programmer can use to make FORTRAN programs more efficient. AO

**Computer Science, P.** Lecture Notes in Computer Science-163: VLSI Engineering: Beyond Software Engineering. Ed: Tosiyasu L. Kunii. Springer-Verlag, 1984, vii + 308 pp, $14 (P). [ISBN: 0-387-70002-1] Focuses on VLSI-related engineering problems as they are used in the production of application-oriented computers. A collection of 14 papers presented at a symposium on VLSI engineering held in Japan in October, 1982. The papers are technical research papers addressing research issues in the manufacture of high density VLSI circuits. MS

**Computer Science, P.** On Conceptual Modelling: Perspectives from Artificial Intelligence, Databases, and Programming Languages. Ed: Michael L. Brodie, John Mylopoulos, Joachim W. Schmidt. Topics in Inform. Sys. Springer-Verlag, 1984, xi + 510 pp, $28. [ISBN: 0-387-90842-0] Many advanced areas of computer science, including artificial intelligence, database design, and programming, must build very high-level abstractions which are independent of a particular problem or computer system. Examples include knowledge representation techniques or semantic networks. This text tries to tie together the ideas of abstraction under the title conceptual modelling. It contains nineteen advanced papers on this subject including discussions and analysis of the subject matter. MS

**Control Theory, T(16-17: 1, 2), P.** Adaptive Filtering Prediction and Control. Graham C. Goodwin, Kwai Sang Sin. Inform. & System Sci. Ser. Prentice-Hall, 1984, xii + 540 pp, $41.95. [ISBN: 0-13-004069-X] A textbook for advanced undergraduate or graduate students that incorporates many recent research results. Both deterministic and stochastic systems are covered. AO

**Control Theory, T(17-18: 1).** Kalman Filtering Theory. A.V. Balakrishnan. Univ. Ser. in Modern Engin. Optimization Software (Distr: Springer-Verlag), 1984, xii + 222 pp, $24 (P). [ISBN: 0-911575-26-X] A textbook for use in a one-quarter or one-semester graduate course. Focuses on those aspects of the subject that can be given a firm mathematical foundation. AO

**Systems Theory, P.** Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory. Harold J. Kushner. Ser. in Signal Processing, Optimization, & Control. MIT Pr, 1984, xvii + 269 pp, $40. [ISBN: 0-262-11090-3] A leading researcher in the area of stochastic processes as applied to control and communication theory explains how known techniques can be used more directly in certain applications, and presents some new methods. AWR

**Systems Theory, P*.** Interdisciplinary Mathematics, V. XXII: Topics in the Geometric Theory of Linear Systems. Robert Hermann. Math Sci Pr, 1984, xvi + 291 pp, $50. [ISBN: 0-915692-35-X] First volume in a projected series "aimed at developing new geometric methodology to be used in engineering and the physical sciences." Directions of future volumes are indicated in author's collection of papers in the Journal of Mathematical Physics since 1980. Forward on "Mathematical Engineering: Problems and Opportunities" is must reading even if one does not agree with author's point of view. JK

**Systems Theory, P.** Stochastic Phenomena and Chaotic Behaviour in Complex Systems. Ed: P. Schuster. Ser. in Synergetics, V. 21. Springer-Verlag, 1984, viii + 271 pp, $31.50. [ISBN: 0-387-13194-9] Collection of the invited contributions of 20 participants in an interdisciplinary workshop of the UNESCO working group on systems analysis. AWR

**Applications, S(16-18), P, L*.** Mathematical Models of Morphogenesis. René Thom. Transl: W.M. Brookes, D. Rand. Ser. in Math. & Its Applic. Halsted Pr, 1983, 305 pp, $59.95. [ISBN: 0-470-27499-9] Translation of the second (1980) French edition, a collection of Thom's papers on morphogenesis dating from 1967 to 1981. Topics range from physics to biology, from linguistics to semiotics; all seek to provide "speculative and philosophical" bases for catastrophe theory as a "stimulus to imagination." LAS

**Applications, P.** Transactions of the First Army Conference on Applied Mathematics and Computing. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC 27709), 1984, xxii + 925 pp, (P). Papers from a May 1983 conference held at George Washington University, combining two symposia--the Conference of Army Mathematicians and the Numerical Analysis and Computers Conference. Over 50 papers on a wide variety of applied topics. LAS

**Applications (Architecture), T(15-17), S, L.** Incidence and Symmetry in Design and Architecture. Jenny A. Baglivo, Jack E. Graver. Urban & Architectural Stud., V. 7. Cambridge U Pr, 1983, xi + 306 pp, $54.50; $15.95 (P). [ISBN: 0-521-23043-8; 0-521-29784-2] Applications of graph theory (e.g., routing, bracing, planning) and transformation geometry (e.g., symmetries of the plane) to architecture. Topics are geometrical (topological) in nature with many excellent illustrations and diagrams. Exercises involve computation, simple theory, as well as paper folding and mirror experiments. LCL

**Applications (Artificial Intelligence), T?(17: 1), P.** Automated Reasoning: Introduction and Applications. Larry Wos, et al. Prentice-Hall, 1984, xiv + 482 pp, $28.95. [ISBN: 0-13-054453-1] Introduction to the theory and practice of automated logical reasoning programming using resolution-based inference rules and strategies. Theoretical foundations, language of clauses, representations, and applications to logic circuit design and validation, research in mathematics and logic, program verification, and expert systems. RM

**Applications (Biology), S(17), P, L.** Graphs and Genes. B.G. Mirkin, S.N. Rodin. Transl: H. Lynn Beus. Biomath., V. 11. Springer-Verlag, 1984, xiv + 197 pp, $27. [ISBN: 0-387-12657-0] Graph theory is applied to the analysis of gene structure, gene semantics and gene evolution. A true melding of mathematical and biological research. Includes a list of references. CEC

**Applications (Biology), P.** The Fluid Mechanics of Large Blood Vessels. T.J. Pedley. Mono. on Mechanics & Appl. Math. Cambridge U Pr, 1980, xv + 446 pp, $97.50. [ISBN: 0-521-22626-0] Aimed at mathematicians interested in cardiovascular fluid mechanics. Provides a comprehensive overview of what is known empirically and of the mathematical understanding of the observed phenomena. Much of the book describes areas still being actively researched. AO

**Applications (Biology), T(14-15: 1), S, L.** Demography Through Problems. Nathan Keyfitz, John A. Beekman. Problem Books in Math. Springer-Verlag, 1984, viii + 141 pp, $28. [ISBN: 0-387-90836-6] Six chapters treat aspects of population theory: age-independent populations, life tables, projection and forecasting, etc. Each chapter contains a brief, intuitive introduction to mathematical concepts, a large collection of problems, and complete or sketched solutions. Assumes a few facts about matrices and a thorough understanding of elementary calculus. PZ

**Applications (Economics), P.** Lecture Notes in Economics and Mathematical Systems-223: Market Demand: An Analysis of Large Economies with Non-Convex Preferences. Walter Trockel. Springer-Verlag, 1984, viii + 205 pp, $14 (P). [ISBN: 0-387-12881-6] Comprehensive treatment of the existing literature on this subject; special emphasis on the "ergodic" approach to smoothing demand by aggregation. JRG

**Applications (Economics), T(14-16: 1), S, L.** An Introduction to Mathematical Models in Economic Dynamics. David L. Clements. Polygonal, 1984, $18.95. [ISBN: 0-936428-07-4] Introduction to mathematical economics. Begins with linear one-sector continuous and discrete models, proceeds to

generalizations: nonlinear and macroeconomic models, trade cycles, stabilization, growth models. Assumes a good mastery of elementary calculus and differential equations--the necessary difference equation material is covered in an appendix. Interesting, attractive application of discrete and continuous mathematics. PZ

**Applications (Engineering), P.** Problems and Methods of Optimal Structural Design. N.V. Banichuk. Transl: Vadim Komkov, Edward J. Haug. Math. Concepts & Methods in Sci. & Eng., V. 26. Plenum Pr, 1983, xxi + 313 pp, $42.50. [ISBN: 0-306-41284-5] An exposition of recent work done by Soviet mathematicians on problems of optimum shape and structural properties of elastic bodies subjected to external forces. Includes a bibliography of Western work in this area. AWR

**Applications (Engineering), P.** Proceedings of Symposium on Finite Element Method. Gordon & Breach, 1982, 478 pp, $92.50. [ISBN: 0-677-3102-0X]

**Applications (Information Theory), T(15: 1), S, P, L.** Probability and Information. A.M. and I.M. Yaglom. Theory & Decision Lib., V. 35. D Reidel, 1983, xx + 421 pp, $69. [ISBN: 90-277-1522-X] The first English edition of this widely read book which originally appeared in 1957. Self-contained, extremely readable and a substantial list of references. No exercises. Includes introductions to probability theory, entropy and information along with applications. CEC

**Applications (Medicine), P.** Lecture Notes in Medical Informatics-23: Selected Topics in Image Science. Ed: O. Malcioglu, Z.-H. Cho. Springer-Verlag, 1984, ix + 308 pp, $25.50. [ISBN: 0-387-12898-0] Papers on diagnostic imaging. Part one concerns algorithms and detectors. Part two deals with applications in emission tomography, digital radiography, ultrasound and nuclear magnetic resonance imaging. RJA

**Applications (Meteorology), P.** Problems and Prospects in Long and Medium Range Weather Forecasting. Ed: D.M. Burridge, E. Källén. Springer-Verlag, 1984, xii + 274 pp, $17.50 (P). [ISBN: 0-387-12827-1] Pioneering work by E. Lorenz shows that the atmosphere is inherently unstable: "Even if we have an almost perfect model with almost perfect initial data, we will never be able to make an accurate weather prediction more than a few weeks ahead." Papers in this volume, from a September 1981 conference in Reading, England, explore the practical consequences of this theoretical limit, and are led off by a survey by Lorenz. LAS

**Applications (Physics), P.** Mechanics Boundary Layers and Function Spaces. Diarmuid Ó. Mathúna. Dublin Inst for Adv Stud, 1984, 216 pp, (P). Monograph on theory of thin elastic plates and shells. The main goal is to make precise features of the passage from the three-dimensional to the simpler two-dimensional theory. It is achieved by a suitable integration in the thickness coordinate of the three-dimensional system equations. PZ

**Applications (Physics), P.** Theory of Dispersed Multiphase Flow. Ed: Richard E. Meyer. Academic Pr, 1983, ix + 388 pp, $28. [ISBN: 0-12-493120-0] Proceedings of an advanced seminar on the motion of multiphase fluids held in the Mathematics Research Center in Madison, Wisconsin in May 1982. Topics range from aerosols to ice crystals, from long chain polymers in liquids to dilute suspension of particles in the atmosphere. LAS

**Applications (Physics), T(18), P*.** Symplectic Techniques in Physics. Victor Guillemin, Shlomo Sternberg. Cambridge U Pr, 1984, xi + 468 pp, $49.50. [ISBN: 0-521-24866-3] An introduction to symplectic geometry from a modern point of view. Applications in physics are an integral part of the presentation. AO

**Applications (Physics), P.** Stochastic Quantum Mechanics and Quantum Spacetime: A Consistent Unification of Relativity and Quantum Theory Based on Stochastic Spaces. Eduard Prugovecki. Fundamental Theories of Physics. D Reidel, 1984, xxi + 302 pp, $48.50. [ISBN: 90-277-1617-X] The author proposes a unification of relativity with quantum mechanics using stochastic phase spaces and stochastic geometries. AO

**Applications (Psychology), P*.** The Stochastic Modeling of Elementary Psychological Processes. James T. Townsend, F. Gregory Ashby. Cambridge U Pr, 1983, xix + 501 pp, $69.50. [ISBN: 0-521-24181-2] Major goals are to introduce the fundamentals of probabilistic modeling of certain simple cognitive processes and to develop mathematical theories encompassing broad classes of models. Good set of references. RSK

## Reviewers

RJA: Richard J. Allen, St. Olaf; MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; PB: Peder Bolstad, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

A great algebraist turned super administrator. (See p. 61.)

$$= 4\int_{\zeta}^{3}(3-t)^{1/2}(t-1)^{1/2}\frac{t\,dt}{\sqrt{t^2-\zeta^2}},$$

where we used the substitution $\sqrt{x^2+\zeta^2}=t$. In particular, when slicing $T$ with $\pi$, the planar surface area $S_p$ is equal to $S(1)$ which is easily evaluated, viz.,

$$S_p = 4\int_1^3 t(3-t)^{1/2}(t+1)^{-1/2}\,dt = -2(3-t)^{3/2}(t+1)^{1/2}\Big|_1^3 = 8. \tag{10}$$

The volume $V$ of the small region obtained by slicing $T$ with $\pi$ is given by

$$V = \int_1^3 S(\zeta)\,d\zeta = 4\int_1^3 d\zeta\int_{\zeta}^{3}(3-t)^{1/2}(t-1)^{1/2}\frac{t\,dt}{\sqrt{t^2-\zeta^2}} \tag{11}$$

$$= 4\int_1^3 t(3-t)^{1/2}(t-1)^{1/2}\arccos\left(\frac{1}{t}\right)dt.$$

In the latter integral we substitute $t = 2+\cos\phi$, and the result is written as

$$V = 4\int_0^{\pi}(2\sin^2\phi - 1 + \cos\phi\sin^2\phi)\arccos\left(\frac{1}{2+\cos\phi}\right)d\phi + 4\int_0^{\pi}\arccos\left(\frac{1}{2+\cos\phi}\right)d\phi. \tag{12}$$

Here, the first integral, to be denoted by $V_1$, is reduced through an integration by parts, yielding

$$V_1 = 4\int_0^{\pi}\frac{(\frac{1}{3}\sin^2\phi - \cos\phi)\sin^2\phi}{(2+\cos\phi)\sqrt{(1+\cos\phi)(3+\cos\phi)}}\,d\phi. \tag{13}$$

The integrand in (13) is reduced to a rational function by the standard substitutions $\cos\phi = t$, $[(1-t)/(3+t)]^{1/2} = s$, viz.,

$$V_1 = \frac{4}{3}\int_{-1}^{1}\frac{1-3t-t^2}{2+t}\left(\frac{1-t}{3+t}\right)^{1/2}dt \tag{14}$$

$$= \frac{32}{3}\int_0^1\frac{(-3+14s^2+s^4)s^2}{(3-s^2)(1+s^2)^3}\,ds = 4\sqrt{3}\log(2+\sqrt{3}) - \frac{16}{3} - \frac{2\pi}{3}.$$

The second integral in (12) follows from (1) and (7). Thus we find for the volume $V$,

$$V = \frac{3\pi^2}{2} - \frac{2\pi}{3} - \frac{16}{3} + 4\sqrt{3}\log(2+\sqrt{3}) - 2\log^2(2+\sqrt{3}) = 13.03207\ldots, \tag{15}$$

where the numerical value has been checked by numerical integration of (12). The curved surface area $S_c$ of the torus slice is given by

$$S_c = 4\int_0^{\pi}(2+\cos\phi)\arccos\left(\frac{1}{2+\cos\phi}\right)d\phi, \tag{16}$$

which can be evaluated by the same methods as before. Omitting further details we present the final result

$$S_c = 3\pi^2 - 2\pi + 4\sqrt{3}\log(2+\sqrt{3}) - 4\log^2(2+\sqrt{3}) = 25.51226\ldots, \tag{17}$$

where the numerical value has been checked by numerical integration of (16).

---

## ANSWER TO PHOTO ON PAGE 49

Irving Kaplansky is Director of the Mathematical Sciences Research Institute, and, as of January 1, 1985, President of the American Mathematical Society.

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## BY ALL MEANS

FRANK BURK
*Department of Mathematics, California State University, Chico, CA 95926*

It is easily shown that

$$\frac{\sinh x}{\sqrt{\sinh^2 x + \cosh^2 x}} < \tanh x < x < \sinh x < \frac{1}{2}\sinh 2x, \quad x > 0. \tag{1}$$

If we take reciprocals in (1) and reverse inequalities, let $x = \ln\sqrt{b/a}$, where $0 < a < b$, and multiply by $\frac{1}{2}(b-a)$, we have

$$\frac{2}{\frac{1}{a}+\frac{1}{b}} < \sqrt{ab} < \frac{b-a}{\ln b - \ln a} < \frac{a+b}{2} < \sqrt{\frac{a^2+b^2}{2}}, \tag{2}$$

i.e., respectively, the harmonic mean, geometric mean, logarithmic mean, arithmetic mean, and the root-mean-square.

## A PROBABILISTIC PROOF OF A THEOREM OF SCHUR*

INGRAM OLKIN
*Department of Statistics, Stanford University, Stanford, CA 94305*

Proofs that an $n \times n$ symmetric matrix $A$ is positive definite (or semidefinite) are usually based on showing that the quadratic form $xAx'$ or the eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$ are positive (nonnegative). We propose another procedure based on the following probabilistic fact: An $n \times n$ matrix $\Sigma$ is positive semidefinite if and only if it is the covariance matrix of a random vector $X = (X_1, \ldots, X_n)$.

If $\Sigma$ is positive semidefinite, then there exist Gaussian random variables having $\Sigma$ as its covariance matrix (see, e.g., [2]). If $X$ is a random vector with mean vector $\mu$, then the covariance matrix $\Sigma$ is $\Sigma = E(x-\mu)'(X-\mu)$, where $E$ is the expectation operator, and hence $\Sigma$ is positive semidefinite. (Since covariances are location invariant we can assume that the means are zero.)

We illustrate the use of this fact to prove the following theorem of Schur (see, e.g., [1], p. 95):

THEOREM. *If $A = (a_{ij})$, $B = (b_{ij})$ are positive semidefinite, then $C = (a_{ij}b_{ij})$ is positive semidefinite.*

*Probabilistic Proof.* Since $A$ and $B$ are positive semidefinite, there exist independent random vectors $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ with zero means and covariance matrices $A$ and $B$, respectively. Define $Z_i = X_i Y_i$, $i = 1, \ldots, n$. Then

$$\begin{aligned} c_{ij} \equiv \mathrm{Cov}(Z_i, Z_j) &= E(X_i Y_i)(X_j Y_j) = (EX_i X_j)(EY_i Y_j) \\ &= \mathrm{Cov}(X_i, X_j)\mathrm{Cov}(Y_i, Y_j) = a_{ij} b_{ij}, \end{aligned}$$

so that $C = (c_{ij})$ is positive definite.

*Supported in part by the National Science Foundation.

This technique can also be used to prove the following theorem about Kronecker products. (See, e.g., [**1**], pp. 235–243.)

THEOREM. *If $A = (a_{ij})$ and $B = (b_{ij})$ are positive semidefinite, then the Kronecker product $A \otimes B = (a_{ij}B)$ is positive semidefinite.*

As a final demonstration of these ideas, we show that the symmetric $n \times n$ matrices

$$A = \begin{bmatrix} n & n-1 & n-2 & \cdots & 1 \\ & 2(n-1) & 2(n-2) & \cdots & 2 \\ & & 3(n-2) & \cdots & 3 \\ & \vdots & \vdots & \vdots & \\ & & & & n \end{bmatrix},$$

$$B = \begin{bmatrix} a_1^2 & a_1^2 & a_1^2 & \cdots & a_1^2 \\ & a_1^2 + a_2^2 & a^2 + a_2^2 & \cdots & a_1^2 + a_2^2 \\ & & a_1^2 + a_2^2 + a_3^2 & \cdots & a_1^2 + a_2^2 + a_3^2 \\ & & \cdots & \cdots & \\ & & \cdots & \cdots & \\ & & & & a_1^2 + \cdots + a_n^2 \end{bmatrix},$$

are positive semidefinite by showing that they are covariance matrices.

(i) If $X_1, \dots, X_n$ are independent random variables from a uniform distribution on $[0,1]$ and $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ are the order statistics, then, except for a constant, the matrix $A$ is the covariance matrix of $(X_{(1)}, \dots, X_{(n)})$.

(ii) The matrix $B$ is the covariance matrix of random variables

$$Z_1 = X_1,\ Z_2 = X_1 + X_2, \dots, Z_n = X_1 + \cdots + X_n,$$

where $X_1, \dots, X_n$ are independent random variables with zero means and variances $a_1^2, \dots, a_n^2$, respectively.


### References

**1.** H. Bellman, Introduction to Matrix Analysis, 2nd ed, McGraw-Hill, New York, 1970.
**2.** H. Cramér, Mathematical Methods in Statistics, Princeton University Press, Princeton, 1946.


## ENUMERATION IN MUSIC THEORY


DAVID L. REINER
*Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439*


The calculation of the number of chords and tone rows using Pólya's Theorem and Burnside's Lemma can add a little variety to the applications of these theorems usually given in combinatorics and algebra texts.

The *n-scale* is taken to be equal (well) tempered and consists of the integers from 0 to $n - 1$. We equate octave notes, so our scale is mathematically $Z_n$ with addition. For example, ordinary Western music has $n = 12$, Debussy used a whole tone scale with $n = 6$, and Ives used a quarter tone scale with $n = 24$. (Of course, in terms of frequencies of pitches, the physical scale is a multiplicative group, but this makes no difference here.)

A *k-chord* in the *n*-scale is an equivalence class (defined in Section 1) of subsets of $k$ elements each of $Z_n$, and an *n-tone row* is an equivalence class (defined in Section 2) of permutations in $S_n$.

**1. Chords.** The equivalence relation is induced by one of two permutation groups. We can use the transpositions (not to be confused with permutations which exchange two elements) $T^i: Z_n \to Z_n: a \mapsto i + a(\text{mod}\, n)$; a jazz guitarist generally uses chords which can be transposed easily so not so many fingerings must be memorized. Or we can use the larger group of transpositions (as above) and the inversion $I: a \to -a(\text{mod}\, n)$. A typical group element looks like $T^iI: a \mapsto i - a(\text{mod}\, n)$. This use of the word "inversion" is standard in serial music and must be distinguished from the usual meaning of octave transposition of one or more notes of a chord. For example, under the larger group, a $C$ major chord is equivalent to a $C$ minor chord: $\{C, E, G\} = \{0,4,7\} \sim \{0,3,7\} = \{C, E\flat, G\}$, the $\sim$ being via $T^7I$.

It is easy to see that these groups are the cyclic groups $C_n$ and the dihedral groups $D_n$, respectively, and we are thus dealing with what is known as the two-color necklace problems, either one-sided or two-sided [**3**, p. 162]. That is, we have a circular necklace of notes to choose from, and we choose the ones in the chord by coloring them one color and the rest another color.

Pólya's Theorem is the appropriate tool for this situation: a group acting on a domain set $D$, inducing a group action on the set of functions into some range set $R$. It suffices to let $G$ be a group of permutations of $D$. $G$ then acts on the set of all functions $R^D$ in a natural way, inducing an equivalence relation.

To keep track of what's important in such a problem, we use enumerators, polynomials or formal power series whose coefficients are the numbers of objects of specified types. The enumerator of the range here can be whatever we like, say $\Sigma_{r \in R} w(r)$, where $w$, the "weight," is any function from $R$ to a commutative ring containing the rationals. The weight of a function from $D$ to $R$ is then defined to be the product of the weights of its image values, and the enumerator or inventory of a set of functions is the sum of the weights of the individual functions. Usually this set of functions is a set of equivalence class representatives when we have an equivalence relation on $R^D$.

In our situation of chords, $R = \{0,1\}$, $w(0) = 1$, $w(1) = x$, and the weight of a function from $D$ to this $R$ is $x^k$, where $k$ is just the size of the subset of $D$ mapped to 1. This is how we enumerate subsets or chords.

The importance of Pólya's theorem is that it relates the inventory of equivalence classes of functions to the "store enumerator," the enumerator of the range, via the cycle index $P_G$ of the group $G$ acting on $D$, which is defined by

$$P_G(t_1, t_2, \dots) = \frac{1}{|G|} \sum_{g \in G} t_1^{a_1} t_2^{a_2} \cdots,$$

where $a_i$ = the number of $i$-cycles in $g$.

Pólya's Enumeration Theorem [**2**, p. 148] states that the enumerator or inventory of equivalence classes under $G$ of functions in $R^D$ is

$$P_G\Big(\sum w(r), \sum w(r)^2, \dots, \sum w(r)^k, \dots\Big).$$

The cycle indices of $C_n$ and $D_n$ are well known [**3**, pp. 149–150]:

$$P_{C_n}(t_1, t_2, \dots, t_n) = \frac{1}{n} \sum_{j|n} \phi(j)\, t_j^{n/j} = \frac{1}{n}\Phi,$$

and

$$P_{D_n}(t_1, t_2, \dots, t_n) \begin{cases} = \dfrac{1}{2n}\left[\Phi + n t_1 t_2^{(n-1)/2}\right], & \text{if } n \text{ is odd,} \\ = \dfrac{1}{2n}\left[\Phi + \dfrac{n}{2} t_1^{(n/2)-1}\left(t_1^2 + t_2\right)\right], & \text{if } n \text{ is even.} \end{cases}$$

Here we use $\phi(j)$, which is the Euler phi-function, the number of positive integers less than $j$ which are relatively prime to $j$, and $\Phi$, which is short for the sum in the first equation.

Our enumerator is just $1 + x$, and $t_j = 1 + x^j$. Hence the number of $k$-chords is the coefficient of $x^k$ in $P(1 + x, 1 + x^2, \ldots, 1 + x^n)$. For $C_n$, a little rearrangement produces

$$\#\text{ of } k\text{-chords} = \frac{1}{n} \sum_{j|(n,k)} \phi(j) \binom{n/j}{k/j} = \frac{1}{n} \Phi_n(k),$$

where $\Phi_n(k)$ is short for the summation. For $D_n$, the additional term gives

$$\#\text{ of } k\text{-chords} = \begin{cases} \dfrac{1}{2n}\left[\Phi_n(k) + n\dbinom{(n-1)/2}{[k/2]}\right], & \text{if } n \text{ is odd,} \\ \dfrac{1}{2n}\left[\Phi_n(k) + n\dbinom{n/2}{k/2}\right], & \text{if } n \text{ is even and } k \text{ is even,} \\ \dfrac{1}{2n}\left[\Phi_n(k) + n\dbinom{(n/2)-1}{[k/2]}\right], & \text{if } n \text{ is even and } k \text{ is odd.} \end{cases}$$

The situation in twelve-tone music is $n = 12$ with dihedral symmetry, and we obtain the following table.

| $k =$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| number | 1 | 1 | 6 | 12 | 29 | 38 | 50 | 38 | 29 | 12 | 6 | 1 | 1 |

Thus, there are 50 hexachords in twelve-tone music and only 6 intervals, a fact well known in music theory, [**1**].

**2. Tone rows.** Here our equivalence classes are induced by the group generated by

transposition $T: S_n \to S_n : (a_1, \ldots, a_n) \mapsto (a_1 + 1, \ldots, a_n + 1) \pmod n$;
inversion $I: (a_1, a_2, \ldots, a_n) \to (a_1, 2a_1 - a_2, \ldots, 2a_1 - a_n) \pmod n$;
retrogradation $R: (a_1, \ldots, a_n) \to (a_n, \ldots, a_1)$.

This group is not well known; however the entire structure of the group is not really needed. We can take a cue from music theorists, who regard transposition as such a basic transformation that they don't work with all permutations, but work with equivalence classes of permutations under transposition, those beginning with 0 being regarded as class representatives. Hence our set is now the set of $(n-1)!$ permutations of $\{1, \ldots, n-1\}$ with a prefix of 0, and our group is generated by $R$ and $I$. On this new set, $RI = IR$, so we have the Klein four group $V$ acting.

Burnside's Lemma is the appropriate tool in this setting.

Burnside's Lemma [**2**, p. 136] states that for a group $G$ acting on $D$, the number of equivalence classes is

$$\frac{1}{|G|} \sum_{g \in G} (\#\ \textit{of elements of } D \textit{ fixed by } g).$$

The computation proceeds as follows (see the appendix for details). In order to avoid triviality, we assume that $n \geqslant 3$.

| Group Element | Number Fixed | |
|---|---|---|
| | $n$ odd | $n$ even |
| $e$ | $(n-1)!$ | $(n-1)!$ |
| $I$ | 0 | 0 |
| $R$ | 0 | $(n-2)(n-4)\cdots(2)$ |
| $IR$ | $(n-1)(n-3)\cdots(2)$ | $(n/2)(n-2)(n-4)\cdots(2)$. |

Thus the number of $n$-tone rows is

$$\tfrac{1}{4}[(n-1)!+(n-1)(n-3)\cdots(2)] \qquad \text{if } n \text{ is odd;}$$
$$\tfrac{1}{4}[(n-1)!+(n-2)(n-4)\cdots(2)(1+n/2)] \qquad \text{if } n \text{ is even.}$$

For example, there are 9985920 twelve tone rows, a fact which does not seem to be in the literature.

**Appendix. Computation of numbers of fixed elements.**

1. $I(0,a_2,\dots,a_n)=(0,-a_2,\dots,-a_n)\sim(0,a_2,\dots,a_n)$ implies $a_i\equiv -a_i \pmod n$ for $i=2,\dots,n$ since no transposition is allowed because the first element 0 is fixed. There is at most one nonzero solution to $x\equiv -x \pmod n$, but that is not enough to fill out the permutation.

2. $R(a_1,\dots,a_n)=(a_n,\dots,a_1)\sim(a_1,\dots,a_n)$ implies that a $t$ exists such that $a_1\equiv a_n+t$, $a_2\equiv a_{n-1}+t,\dots \pmod n$. If $n$ is odd, the middle element is fixed and no transposition is allowed: $t=0$. But then $a_n=a_1$, a contradiction. If $n$ is even, the first and last congruences imply that $2t\equiv 0$; hence, $t=0$ or $t=n/2$. The first is impossible just as when $n$ is odd, but the other gives fixed permutations. Since $a_1=0$, $a_n=n/2$. For $a_2$, we can choose any of $n-2$ elements, and this determines $a_{n-1}$. For $a_3$, we have $n-4$ choices, etc.

3. $IR(a_1,\dots,a_n)=(-a_n,\dots,-a_1)\sim(a_1,\dots,a_n)$ implies that a $t$ exists such that $a_1+a_n\equiv t$, $a_2+a_{n-1}=t,\dots$. The last congruence for $n$ odd is $2a_{(n+1)/2}\equiv t \pmod n$. Clearly it is not important that we fix the first element as 0; we could fix $a_{(n+1)/2}$ as 0 and obtain the same count. Thus we may assume that $t=0$, $a_{(n+1)/2}=0$. This allows $n-1$ choices for $a_1$, with $a_n$ thus determined, $n-3$ choices for $a_2$, etc. For $n$ even, we fix $a_1=0$, $a_n\neq t\neq 0$. A little thought shows that $t$ must be odd in order for us to complete the permutation. If $t=2k$, then there is no mate for $k$ in the permutation. There are thus $n/2$ choices for $t=a_n$. For $a_2$, there are $n-2$ choices, with $a_{n-1}$ determined thereby, etc.

The author thanks Dennis White for helpful comments.

**References**

1. A. Forte, The Structure of Atonal Music, Yale University Press, New Haven, 1973.
2. C. L. Liu, Introduction to Combinatorial Mathematics, McGraw-Hill, New York, 1968.
3. J. Riordan, An Introduction to Combinatorial Mathematics, Wiley, New York, 1967.

## BIJECTING EULER'S PARTITIONS-RECURRENCE

DAVID M. BRESSOUD
*Department of Mathematics, Pennsylvania State University, University Park, PA 16802*

DORON ZEILBERGER
*Department of Mathematics and Computer Science, Drexel University, Philadelphia, PA 19104*

A partition of an integer $n$ is a nonincreasing sequence of positive integers $\lambda(1)\geqslant\lambda(2)\geqslant\cdots\geqslant\lambda(t)>0$, such that $\lambda(1)+\cdots+\lambda(t)=n$. The set of partitions of $n$ is denoted $\mathrm{Par}(n)$ and its cardinality $|\mathrm{Par}(n)|$ is written $p(n)$. For example,

$$\mathrm{Par}(5)=\{5;4,1;3,2;3,1,1;2,2,1;2,1,1,1;1,1,1,1,1\} \quad\text{and}\quad p(5)=7.$$

There is no closed form formula for $p(n)$ but Euler ([**1**], p. 12) gave a very efficient way for compiling a table of $p(n)$ by proving the recurrence

$$(1)\qquad \sum_{j\text{ even}} p(n-a(j)) = \sum_{j\text{ odd}} p(n-a(j)), \quad \text{where } a(j)=(3j^2+j)/2.$$

Euler used generating functions to prove this formula. Garsia and Milne [**2**] gave a very nice

bijective proof of (1), utilizing their Involution Principle. We are going to give another bijective proof which does not require any iterations and is very simple. Indeed,

$$\phi\colon \bigcup_{j \text{ even}} \mathrm{Par}(n - a(j)) \leftrightarrow \bigcup_{j \text{ odd}} \mathrm{Par}(n - a(j)),$$

given below does the job.

Let $(\lambda) = (\lambda(1), \ldots, \lambda(t)) \in \mathrm{Par}(n - a(j))$. Then define $\phi$ by

$$\phi((\lambda)) = \begin{cases} (t + 3j - 1, \lambda(1) - 1, \ldots, \lambda(t) - 1) \in \mathrm{Par}(n - a(j-1)), & \text{if } t + 3j \geqslant \lambda(1), \\ (\lambda(2) + 1, \ldots, \lambda(t) + 1, 1, 1, \ldots, 1) \in \mathrm{Par}(n - a(j+1)), & \text{if } t + 3j < \lambda(1), \\ \text{where there are } \lambda(1) - 3j - t - 1 \text{ 1's at the end.} & \end{cases}$$

Note that applying $\phi$ twice yields the identity mapping, thus $\phi = \phi^{-1}$ and $\phi$ is invertible.

EXAMPLE. $n = 21$.

$$\phi(5,5,4,3,2) = (7,4,4,3,2,1).$$

Here $(5,5,4,3,2,) \in \mathrm{Par}(19) = \mathrm{Par}(n - a(1))$ so $j = 1$. The number of parts $t$, is 5 and we have $t + 3j \geqslant \lambda(1)$, since $5 + 3 \geqslant 5$. Now consider $\phi(7,4,4,3,2,1)$; here $j = 0$, $t = 6$, $\lambda(1) = 7$ and $6 + 0 < 7$. Also $\lambda(1) - 3j - t - 1 = 7 - 0 - 6 - 1 = 0$ so we do not add any 1's at the end and $\phi(7,4,4,3,2,1) = (5,5,4,3,2)$.

### References

1. G. E. Andrews, The Theory of Partitions, Addison-Wesley, Reading, MA, 1976.
2. A. M. Garsia and S. C. Milne, A Rogers-Ramanujan bijection, J. Combin. Theory (A), 31 (1981) 289–339.

---

# THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## ON THE CONVOLUTION OF CAUCHY DISTRIBUTIONS

MEYER DWASS
*Department of Mathematics, Northwestern University, Evanston, IL 60201*

The characteristic function of a probability distribution is usually too advanced a topic for a first undergraduate course in mathematical statistics and the more limited moment generating function is often used instead. In teaching the distribution of sums of independent random variables such as normal, gamma, or uniform, I supplement the use of the moment generating function with the convolution formula,

$$f * g(u) = \int_{-\infty}^{\infty} f(x) g(u - x)\, dx. \tag{1}$$

For sums of independent Cauchy random variables the moment generating function does not apply and the use of the convolution formula is difficult. Undoubtedly, it is generally understood that if $f$ and $g$ are Cauchy densities, a partial fraction decomposition of the integrand in (1) should lead to an explicit evaluation of the convolution integral but I do not find the details worked out anywhere. (See comment in Feller, [1], p. 51.) The purpose of this note is to outline

these details as I have presented them in the classroom. The algebra is elementary and well within the scope of an undergraduate course.

The calculation is based on the following relationship:

LEMMA. *For* $a > 0$, *let*

$$f(x,a) = \frac{1}{\pi a\left[1 + (x/a)^2\right]}, \qquad -\infty < x < \infty.$$

*Then*

$$f(x,a)f(u-x,b) = \frac{[a+b-ux/a]f(x,a) + [a+b-u(u-x)/b]f(u-x,b)}{2\pi\left[(a+b)^2+u^2\right]} + \frac{[-a+b+ux/a]f(x,a) + [a-b+u(u-x)/b]f(u-x,b)}{2\pi\left[(a-b)^2+u^2\right]}.$$

From the lemma it follows immediately that

$$\int_{-\infty}^{\infty} f(x,a)f(u-x,b)\,dx = f(u,a+b), \tag{2}$$

since the integral in (2) exists and

$$\int_{-\infty}^{\infty} f(x,a)f(u-x,b)\,dx = \lim_{T\to\infty}\int_{-T}^{T} f(x,a)f(u-x,b)\,dx$$

and

$$\lim_{T\to\infty}\int_{-T}^{T} xf(x,a)\,dx = \lim_{T\to\infty}\int_{-T}^{T}(u-x)f(u-x,b)\,dx = 0.$$

*Proof of Lemma*. For $a > 0$, define

$$h(x,a) = \frac{1}{2\pi a(1+ix/a)},$$

$$h(x,-a) = \frac{1}{2\pi a(1-ix/a)},$$

where $i = \sqrt{-1}$.

The following steps follow by routine algebra and the details are left to the reader:

(a) $f(x,a) = h(x,a) + h(x,-a)$,

(b) $f(x,a)f(u-x,b) = h(x,a)h(u-x,b) + h(x,a)h(u-x,-b) + h(x,-a)h(u-x,b) + h(x,-a)h(u-x,-b)$,

(c) $h(x,a)h(u-x,b) = \dfrac{h(x,a)+h(u-x,b)}{2\pi(a+b+iu)}$,

$$h(x,a)h(u-x,-b) = \frac{-h(x,a)+h(u-x,-b)}{2\pi(a-b+iu)},$$

$$h(x,-a)h(u-x,b) = \frac{-h(x,-a)+h(u-x,b)}{2\pi(a-b-iu)},$$

$$h(x,-a)h(u-x,-b) = \frac{h(x,-a)+h(u-x,-b)}{2\pi(a+b-iu)},$$

(d) $h(x,a)h(u-x,b)+h(x,-a)h(u-x,b)=$
$$\frac{[a+b-ux/a]f(x,a)+[a+b-u(u-x)/b]f(u-x,b)}{2\pi\left[(a+b)^2+u^2\right]},$$
$h(x,a)h(u-x,-b)+h(x,-a)h(u-x,b)=$
$$\frac{[-a+b+ux/a]f(x,a)+[a-b+u(u-x)/b]f(u-x,b)}{2\pi\left[(a-b)^2+u^2\right]}.$$
The lemma now follows by combining the last two equations.

**Reference**

**1.** William Feller, An Introduction to Probability Theory and Its Applications, vol. II, 2nd ed., Wiley, New York, 1971.

---

# PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by May 31, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3069. *Proposed by Zhang Zaiming, Yuxi Teachers' College, Yunan, China.*

The $m\times m$ determinant $I=|a_{rs}|$ has $a_{rs}=\int_0^1 x^{s-1}[F(x)]^{m-r+1}\,dx$, where $F$ is nondecreasing on $[0,1]$. Prove that $I\geqslant 0$.

E 3070. *Proposed by Gérard Letac, Université Paul Sabatier, Toulouse, France.*

Let $a$ and $(t_n)_{n\geqslant 1}$ be strictly positive numbers and let the sequence $(t_n)_{n\geqslant 1}$ be bounded. Prove

that

$$t_1^{-a} + \sum_{n=1}^{\infty} t_1 t_2 \cdots t_n t_{n+1}^{-a} \geqslant \sum_{n=0}^{\infty} (a/(a+1))^{n-a}.$$

E 3071. *Proposed by K. Satyanarayana, Hyderabad, India.*

If in triangle $ABC$, angle $C >$ angle $B >$ angle $A$, then prove that $I$ lies inside triangle $OBH$ where, as usual, $O, I, H$ denote the circumcenter, incenter and orthocenter, respectively, of triangle $ABC$.

E 3072. *Proposed by Louis Funar, Craiova, Rumania.*

Let $P(x) = a_0 + a_1 x + \cdots + a_n x^n$ be a real polynomial of degree $n \geqslant 2$ such that

$$0 < a_0 < -\sum_{k=1}^{[n/2]} \frac{1}{2k+1} a_{2k}.$$

Prove that $P(x)$ has a real zero $r$ such that $|r| < 1$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### A Four-Point Tangency

E 2911 [1981, 705]. *Proposed by Jordi Dou, Barcelona, Spain.*

Let 2 semicircles $AC, CB, \overline{AC} = 3\overline{CB}$, be given ($ACB$ are collinear). Let $a, b$ be the tangents to the given semicircles at $A, B$. Let $\gamma$ be the circle tangent to $a, b$ and to the larger of the given semicircles. Prove $\gamma$, $b$, and the given semicircles have a common tangent circle.

*Solution by William A. Newcomb, Lawrence Livermore National Laboratory, CA.* Let a Cartesian coordinate system be introduced such that the points $A, C, B$ are assigned the respective coordinates (0,0), (0,6), and (0,8). Also, let us use the notation $(\alpha, \beta, r)$ to denote the circle whose equation is $(x-\alpha)^2 + (y-\beta)^2 = r^2$.

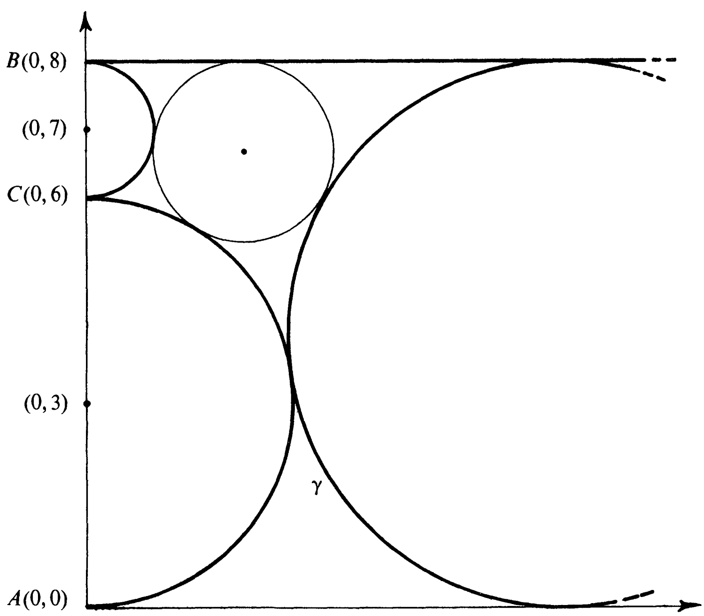



FIG. 1

The criterion for two circles, $(\alpha, \beta, r)$ and $(\alpha', \beta', r')$, to be externally tangent is

$$\left[(\alpha - \alpha')^2 + (\beta - \beta')^2\right]^{1/2} = r + r'.$$

The two given circles are $(0,3,3)$ and $(0,7,1)$, and the circle $\gamma$ is $(4\sqrt{3}, 4, 4)$. The circle $(4\sqrt{3}/3, 20/3, 4/3)$ is externally tangent to all three of these circles, and also tangent to the straight line $y = 8$.

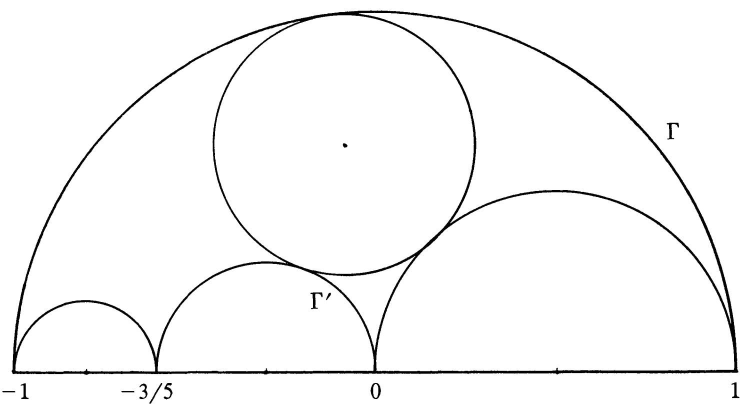



FIG. 2. After inversion with respect to $\Gamma$.

*Solution by I. J. Schoenberg, University of Wisconsin-Madison.* In $\mathbb{C}$ we choose a coordinate system so that $A = +1$, $C = -1$, and therefore $B = -5/3$. If $\alpha < \beta$ are reals, we denote by $(\alpha, \beta)$ the circle of diameter $[\alpha, \beta]$. Performing an inversion with respect to the circle $\Gamma = (-1, 1)$ (see Fig. 2), we find that the tangent $b$ goes over into the circle $\Gamma' = (-3/5, 0)$, and that the problem is equivalent to the statement: *The* (excentric) *ring between the two circles* $\Gamma, \Gamma'$ *contains a* (closed) *Steiner system of six circles which are pairwise tangent and all tangent to* $\Gamma$ *and* $\Gamma'$. To make the ring $\Gamma, \Gamma'$ *concentric*, we perform a real Möbius transformation $M$ such that $M\Gamma = \Gamma$ and $M\Gamma' = (-a, a) = \Gamma^*$. The invariance of the anharmonic ratio shows that we must have $(-1, -a, a, 1) = (-1, -3/5, 0, 1)$ whence $a = 1/3$. Since the circle $(1/3, 1)$ is seen from the origin 0 under an angle of $60°$, the Steiner chain of six circles within the ring between $\Gamma$ and $\Gamma^*$ becomes obvious (see Fig. 3).

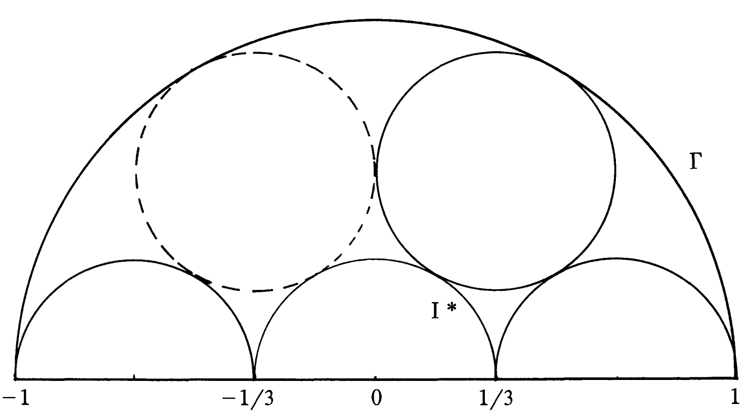



FIG. 3. The sixth circle is now obvious.

Also solved by the proposer.

### Six Concurrent Circles

E 2963 [1982, 594]. *Proposed by Calin P. Popescu, student, Bucharest, Rumania.*

Let $A_1A_2A_3, A_1'A_2'A_3'$ be two equilateral triangles in the plane. Construct circles $\gamma_i$ $[\gamma_i']$ with radii $r_i$ $[r_i']$ and centers $A_i$ $[A_i']$, $i = 1,2,3$ $[i = 1,2,3]$. Suppose further that $r_i$ $[r_i']$ are geometric progressions with ratio a positive integer. When can the six circles be concurrent?

*Solution by L. Kuipers, Switzerland.* We show that three circles centered at the vertices of an equilateral triangle $A_1A_2A_3$ and with radii in geometric progression with integer ratio $m$ are concurrent at a point $P$ only if $m = 1$, in which case $P$ is the centroid of the triangle. Thus the six circles in the problem are concurrent only if the two triangles have a common centroid.

Any point $P$ such that

$$\frac{A_2P}{A_1P} = m = \frac{A_3P}{A_2P}$$

must be a point of intersection of the two Appolonian circles determined by the first and second equations respectively. It is easy to see that these circles do not intersect if $m \geqslant 2$. Hence $m = 1$, in which case the circles degenerate to two medians intersecting at the centroid $P$ of $A_1A_2A_3$.

Also solved by J. Dou (Spain), O. P. Lossers (The Netherlands) and the proposer.

*Editorial note*: The proposer wishes to say that he was inspired by the memory of his friend Stephen Ghiviragă.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by May 31, 1985. The solver's full post-office address should be on each sheet.*

6484. *Proposed by A. V. Nabutovskii, Novosibirsk, U.S.S.R.*

Suppose that $x = \sum_{n=1}^{\infty} p_n/q_n < \infty$, where the $p_n$'s and $q_n$'s are positive integers satisfying the inequality

$$\frac{p_n}{q_n(q_n - 1)} \geqslant \frac{p_{n+1}}{q_{n+1} - 1} \qquad \text{for } n = 1,2,\dots .$$

Let $S$ be the set of indices $n$ for which the inequality is strict. Prove that $x$ is irrational if and only if $S$ is infinite.

6485. *Proposed by Alexandru Lupaş, Sibiu, Rumania.*

Given a function $f \in C[0,1]$, let $(B_nf)(x)$ denote the Bernstein polynomial

$$\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right).$$

If $f \in C^{(2)}[0,1]$, prove that, for $0 \leqslant x \leqslant 1$, $n = 1,2,\dots$,

$$|(B_nf)(x) - (B_{n+1}f)(x)| \leqslant \frac{x(1-x)}{n+1}\left(\frac{1}{3n}\int_0^1 |f''(t)|^2\,dt\right)^{1/2}.$$

6486. *Proposed by Liviu I. Nicolaescu (student), University Ai Cuza Iasi, Rumania.*

Let $S_n$ denote the set of all permutations of $1,2,\dots,n$. For $\sigma, \tau \in S_n$, define $d(\sigma,\tau)$ to be the number of inversions of the permutation $\sigma \cdot \tau^{-1}$. Prove that $d$ is a metric on $S_n$.

6487. *Proposed by Pei Yuan Wu, National Chiao Tung University, Hsinchu, Taiwan, Republic of China.*

Let $T$ be a contraction (i.e., $\|T\| \leqslant 1$) on a Hilbert space and, for $n = 1, 2, \ldots,$ let

$$K_n = 1 + \sum_{k=1}^{n}\left(1 - \frac{k}{n+1}\right)(T^k + T^{*k})$$

be its "Fejér kernel". Prove by an elementary method (i.e., without the use of spectral theory or dilation theory) that $K_n \geqslant 0$ for $n = 1, 2, \ldots$ .

## SOLUTIONS OF ADVANCED PROBLEMS

### An Exponential Diophantine Equation

6411 [1982, 788]. *Proposed by Leo J. Alex, SUNY, Oneonta.*

Find all solutions in integers $a, b, c, d$ to the equation $1 + 3^a = 5^b + 3^c5^d$.

*Solution by the proposer.* We show that the only solutions in integers are $(a, b, c, d) = (t, 0, t, 0)$, $(2, 1, 0, 1)$ or $(3, 2, 1, 0)$. The key to proving this is a result which follows from work of R. J. Stroeker and R. Tijdeman, [*Diophantine Equations, Computational Methods in Number Theory*, MC Tract 154(Part I), MC Tract 155(Part II), (1983), pp. 321–369]:

LEMMA. *If* $0 < |3^u - 5^v| < 3^{u/2}$, *where u and v are integers, then* $(u, v) = (3, 2)$.

We now consider the given equation. If $b = 0$, then $a = c$, $d = 0$; i.e., $(a, b, c, d) = (t, 0, t, 0)$. Thus we may assume that $b > 0$. Then $a > c$ and so $5^b \equiv 1 \pmod{3^c}$. Since $5^2 \not\equiv 1 \pmod{3^2}$, 5 is a primitive root mod $3^c$. It follows that $2 \cdot 3^{c-1} | b$ when $c > 0$, and hence that $3^c \leqslant 3b/2$. We distinguish the Cases (i) $b \geqslant d$ and (ii) $b < d$.

In Case (i) $3^a \equiv -1 \pmod{5^d}$ and so, since 3 is a primitive root mod $5^d$, $2 \cdot 5^{d-1} | a$ when $d > 0$. Hence, $5^d \leqslant 5a/2$ and so $0 < 3^a - 5^b \leqslant 15ab/4$. Consequently, $b \leqslant a \log 3/\log 5$ and $0 < 3^a - 5^b \leqslant 2 \cdot 6a^2$. By the lemma we have $|3^a - 5^b| > 3^{a/2}$ if $(a, b) \neq (3, 2)$. Hence, $2 \cdot 6a^2 > 3^{a/2}$, and so $a \leqslant 10, b \leqslant 6, c \leqslant 2, d \leqslant 2$. It follows easily that $(a, b, c, d) = (2, 1, 0, 1)$ or $(3, 2, 1, 0)$.

In Case (ii) we have $3^a \equiv -1 \pmod{5^b}$, whence $5^b \leqslant 5a/2$. Thus $0 < 3^a - 3^c5^d < 5^b \leqslant 5a/2$, and hence $0 < 3^{a-c} - 5^d \leqslant 5a3^{-c}/2$. Suppose first $(a - c, d) \neq (3, 2)$, so that, by the lemma, $|3^{a-c} - 5^d| > 3^{(a-c)/2}$. Then $3^{a/2} \leqslant 5a/2$, and so $a \leqslant 4$, $b = 1$, $d \geqslant 2$. Since $5 | 3^a + 1$, we obtain $a = 2$, a contradiction. Finally, suppose $(a - c, d) = (3, 2)$. Then $b = 1$, $c = 0$, and thus $3^a = 2a$, again a contradiction. Therefore there are no solutions in Case (ii).

Also solved by I. A. Sakmar (Canada), and partially solved by Mihàly Bencze (Rumania).

### Partition of Integers

6428 [1983, 338]. *Proposed by J. M. Steele, Princeton University.*

Let $\varepsilon$ be a separation of $\{1, 2, \ldots, 2n\}$ into two disjoint classes $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots b_n$. Let $M_k$ denote the number of solutions of $a_i - b_j = k$, $-2n < k < 2n$, and put $M(n) = \min_\varepsilon \max_k M_k$. Selfridge, Motzkin, and Ralston showed $\liminf n^{-1}M(n) \leqslant .4$, and Moser showed $\limsup n^{-1}M(n) \geqslant .357$. Prove that the limit $\lim n^{-1}M(n)$ actually exists.

*Solution by the proposer.* Let $M^* = \liminf n^{-1}M(n)$, $\delta > 0$, and choose an $n$ such that $n^{-1}M(n) \leqslant M^* + \delta$. If the optimal separation $\varepsilon$ for this $n$ is given by $(a_i)$, $(b_j)$, we consider the separation $\varepsilon'$ of $\{1, 2, \ldots, 2nr\}$ given by $(ra_i + h)_{i=1}^n$, $0 \leqslant h < r$ and $(rb_j + h')_{j=1}^n$, $0 \leqslant h' < r$. The number of elements of this separation with distance $k$ is equal to the number of solutions of $(ra_i + h) - (rb_j + h') = k$, and this is easily seen to be bounded by $rM(n)$.

This proves $M(rn) \leqslant rM(n)$; hence $\limsup_{n\to\infty}(rn)^{-1}M(rn) \leqslant M^* + \delta$. Since $M(rn+h) \leqslant M(rn) + h$ we also have $\limsup n^{-1}M(n) \leqslant M^* + \delta$, which suffices to prove the claim.

Also solved by Tim Keller.

**Differences Between Consecutive Primes**

6429 [1983, 338]. *Proposed by B. Powell, Kirkland, WA.*

Prove that for any positive integer $M$ there exists an even positive integer $n_0 = n_0(M)$ such that the number of pairs of consecutive primes $(P_j, P_{j+1})$ that differ by exactly $n_0$ is greater than $M$.

*Solution by Roy O. Davies, The University of Leicester, England.* Suppose the assertion false. Then there is a bound $M$ to the number of times any difference between successive primes appears. In particular, there are at least $(n-2)/M$ distinct differences between successive primes in the sequence $3 = p_2, p_3, \ldots, p_n$, and so

$$p_n - p_2 \geqslant 2 + 4 + \cdots + 2\left[\frac{n-2}{M}\right] \sim \frac{n^2}{M^2}.$$

But this contradicts the elementary estimate $p_n = O(n \log n)$.

Also solved by Richard Beigel, Robert Breusch, Keith Kearnes, O. P. Lossers (Netherlands), Victor Pambuccian (Rumania), University of South Alabama Problem Group, and the proposer.

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Sophie Germain. An Essay in the History of the Theory of Elasticity.* By Louis L. Bucciarelli and Nancy Dworsky. D. Reidel, Boston, 1980. xii +143 pp. \$31.50 (cloth), \$15.75 (paper).

JOSEPH W. DAUBEN
*Herbert H. Lehman College, CUNY, Bedford Park Boulevard West, Bronx, NY 10468*

This is a book whose title should have been reversed, for accuracy's sake, to read: *An Essay in the History of the Theory of Elasticity: the (Minor) Role of Sophie Germain.* Reading the authors' preface, which describes the progress of their research and thinking that eventually led to this brief history, it becomes clear that elasticity is their real interest. Almost accidentally, however, they discovered "that a woman appeared to have made a significant contribution in the field" (page 6), and she subsequently aroused their curiosity. Of course, one understandably suspects why both the authors and their publisher chose to emphasize Germain in the title, for (apart from specialists) who would be likely to read, let alone buy, a history of theories concerning elasticity? The additional story, however, of a French woman born in the revolutionary year 1776, the self-taught product of France's own decade of upheaval, who not only took it upon herself to study mathematics and its physical applications, but to compete for a *prix extraordinaire* offered by the Institut de France, and then actually *win* its gold metal in 1816—might well attract many more readers, certainly those interested in the social history of women, as well as those interested in mathematics and its more esoteric aspects. Unfortunately, in point of fact, Germain was neither a very good mathematician, nor very successful in the applications she tried to make of mathematical techniques in her various studies of elasticity. Nor does she appear to have been an exceptionally interesting personality. Although little in fact is known about her life, what this biography does manage to reveal suggests that Germain was possibly the most lackluster of the

great or near-great women, not only of her own day—consider Madame de Staël, for example—but of the recent history of mathematics as well. Germain's biography lacks any of the social or political interest of Sophie Kovalevskaya's, and she offers none of the originality or depth of mathematical insight that characterize Emmy Noether's work. But the authors of *Sophie Germain* are not responsible for their subject's shortcomings, personality, or for the reclusive and rather prosaic life she seems to have led. Instead, it is necessary to ask what their book sets out to do, and what interests it manages to serve?

Readers of this MONTHLY will naturally be interested primarily in the mathematical aspects of her work, which can be summarized rather briefly. Her most impressive results were in number theory. A limited correspondence with Gauss, which survives in the Universitätsbibliothek, Göttingen, and in the Bibliothèque Nationale, Paris, shows that he was impressed not only with what she had accomplished in this area, but also with the fact that she was a woman. This did not, however, affect the first impressions he had of her work, for when she initiated their correspondence in 1804, she used the nom de plume "Monsieur LeBlanc." Two years later, however, when Napoleonic forces were occupying Prussia, Germain was concerned with Gauss' safety in Brunswick, and used family connections to have a French army officer intercede personally on Gauss' behalf. Later, she wrote to him, explaining that "fearing the ridicule attached to a female scientist," she had "previously taken the name of M. LeBlanc in communicating to you those notes that, no doubt, do not deserve the indulgence with which you have responded." As a result of this revelation, however, Gauss was not only indebted to Germain for her concern, but now was doubly interested in her mathematics and in the fact that she was a woman. "The taste for the abstract sciences in general and, above all, for the mysteries of numbers, is very rare," he responded. "But when a woman, because of her sex, our customs and prejudices, encounters infinitely more obstacles than men in familiarizing herself with their knotty problems, yet overcomes these fetters and penetrates that which is most hidden, she doubtless has the most noble courage, extraordinary talent, and superior genius," (page 25). This was, easily, the most extensive and best-informed praise Germain ever received for any of her scientific work.

In fact, her accomplishments in number theory have immortalized her eponymously. It was her friend and mentor Legendre who published the theorem proving Fermat's Last Theorem for the special cases $n = p - 1$, where $p$ is a prime of the form $8k + 7$, that is today known as Germain's Theorem. As Germain once said of herself, in a letter to Gauss written in 1809, her mind had a "great predilection for arithmetic problems" (page 121).

Unfortunately, she had neither the same innate abilities, nor the opportunities, to perform so well in other branches of mathematics, particularly those that were basic for mathematical physics, and in fact essential for her studies of elasticity. Here, despite the celebrated *prix extraordinaire*, her work was deeply flawed. Even her biographers describe her results variously as "naive and unconvincing" (page 93), "slight and unsuccessful" (page 105), "inadequate and trivial" (page 107). But such negative evaluations concern primarily her difficulties with higher mathematics, especially the calculus and variational techniques.

If Germain's mathematical physics was less than adequate, however, how was she able to have won the *prix extraordinaire* from the French Académie des Sciences? The answer depends upon the state of mathematical physics at the turn of the century, and begins with the dramatic public lectures delivered in Paris on vibrating plates by E. F. F. Chladni in 1802. A metal disc, caused to vibrate, creates a diversity of different geometric patterns, and the challenge (apart from setting up the vibrations, which required considerable skill) was of course to account for the nodal lines mathematically. (See Fig. 1.1 on p. 66.) Napoleon, who was not only a member of the Institut, but in fact of the First Class (for physics and mathematics)—he served as President of both—was so impressed by Chladni's demonstrations that he urged a prize competition be established on the subject. In 1809 the competition was announced, with a prize set at a kilogram of gold (about 3000 francs) for "the development of a mathematical theory of the vibration of elastic surfaces, and a comparison of this theory with experiments" (page 35).

Germain's first paper on the subject, submitted in 1911, was also the *only* entry. (G. Libri, Germain's friend and first biographer, conjectured that potential competitors had been discour-

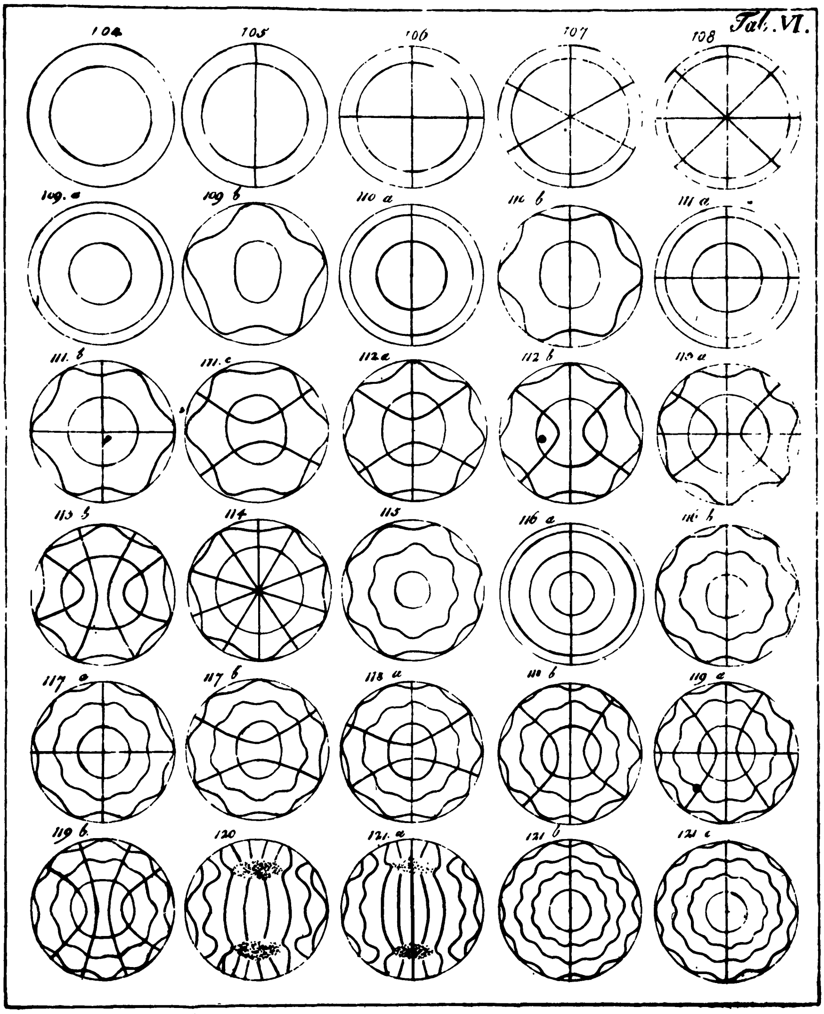


FIG. 1.1. Chladni's vibration modes of circular plates.

aged from working on the problem by Lagrange's pronouncement that the subject was extraordinarily difficult and would require an entirely new kind of analysis.) In fact, Lagrange noted the major difficulty that Germain took as her starting point (never successfully): "We do not even possess the differential equations of movement for this type of vibration in considering these phenomena as nature presents them" (page 43).

Germain began with Euler's treatment of horizontal beams, and attempted to generalize mathematically to the case of vibrating surfaces. Through correspondence (and perhaps actual meetings) with Legendre, she tried to refine this approach, but she was never successful. As Legendre wrote after her paper had failed to win any notice from the Académie, her attempts to introduce double integrals were "nowhere amenable to the substitutions you have made" (page 54).

However, her paper did contain a valuable assumption about the *nature* of the problem that inspired Lagrange to deduce the proper differential equation for vibrating plates. Beginning with the assumption that the elasticity at every point of a disc or plate could be represented by resistances of the form $(1/r) + (1/r')$, where $r$ and $r'$ are the two principal radii of curvature of the deformed plate, Germain had produced an erroneous differential equation.

Lagrange, however, rejected Germain's faulty deductions, and with his own methods reached

the alternative equation:

$$\frac{d^2z}{dt^2} = k^2\left(\frac{d^4z}{dx^4} + 2\frac{d^4z}{dx^2\,dy^2} + \frac{d^4z}{dy^4}\right).$$

Legendre communicated the gist of Lagrange's conclusions to Germain, who thereafter used Lagrange's correct equation for her own subsequent work on elasticity. The prize contest was extended until October of 1813, when she submitted a second entry, again the only one received. This time she presented the equation that Lagrange had deduced in response to her first paper, but she offered little justification for the result. Again, her analysis was so full of errors that the judges were not convinced by her mathematical deductions. As Legendre eventually explained to her in December: "I do not understand the analysis you send me at all; there is certainly an error in the writing or the reasoning, and I am led to believe that you do not have a very clear idea of the operations on double integrals in the calculus of variations" (page 63).

But the second part of her paper aroused greater interest, and was sufficient to bring her at least a "honorable mention." This was awarded for the connection she established between the physical, experimental data and (Lagrange's) correct differential equation.

Germain's confirmation of the applicability of the mathematical interpretation of the phenomena soon inspired Poisson, apparently, to produce an impressive (but incorrect) mathematical theory of elasticity. His work eventually involved both Germain and Navier in a heated polemic with Poisson over priorities and approaches, and this constitutes, easily, the most interesting and impressive parts of the study produced by Bucciarelli and Dworsky. This aspect of their book, reflected in its *subtitle*, deserves careful reading, for it is filled with information about how the theory of elasticity developed in the 19th Century, based upon false assumptions, heated rivalries, erroneous metaphysics, and last, but not least, elaborate but false mathematical models of a particularly interesting set of physical phenomena.

Not satisfied with Germain's second memoir, the Academy extended yet again the prize competition on elasticity, setting a deadline of October, 1815. Germain, not satisfied with her "honorable mention," began work on yet another memoir, this time working wholly in isolation, with no help (or encouragement, it seems) from Legendre or anyone else.

Meanwhile, Poisson read a paper to the Institut in August, 1814, attempting a rigorous derivation of the differential equation for elasticity based on the "molecular hypothesis" (page 75). He produced a nonlinear, more general equation than had Lagrange. Germain, impressed by Poisson's reputation and unable to follow the details of his mathematical arguments, showed in her third memoir that her own hypothesis led to the same equation Poisson had produced! But this result only underscores how ineffectually Germain manipulated, but could hardly justify, the results she obtained.

Germain's third memoir on elasticity finally was awarded the *prix extraordinaire* in 1816 (again it was the sole entry). Although it was "interesting in intent" (faint praise), Bucciarelli and Dworsky admit that it was "fundamentally deficient and wrong in execution" (page 80). Moreover, before Germain had even submitted her paper, Biot had already decided that Poisson's memoir a year earlier had "resolved" the problem of elasticity. So the problem remains, why, at last, in a version hardly better than any of her earlier attempts, should Germain now have been awarded the prize? Was it simply a means of ending the matter once and for all?

Germain, having won the *prix extraordinaire* of the Academy in 1816 (in absentia, for she realized the interest her appearance for the award would create, and wished to avoid any public spectacle), went on to publish her memoir in a further, revised form at her own expense in 1821. Despite revisions and reworkings, Bucciarelli and Dworsky are not hesitant to say that it still revealed "a tenuous grasp on relevant physical concepts and a less than full understanding of the substance and technique of the variational approach of Lagrange" (page 92). Worse still, publication "now displayed her incompetence" publicly (page 93).

The published memoir also made clear her differences with Poisson. She was openly on record, after 1821, of opposing the "molecular mentality" and instead, advanced her own hypothesis for

the first principles of elasticity theory (page 94), hoping that "an informed public may weigh them." Her confidant by now, Lagrange, was immediately worried that this would draw Poisson into open confrontation, and he hoped Germain would not "find occasion to repent of (her) courage," for he doubted that Poisson would be able "to respond with dignity to this civilized assault" (page 96).

Eventually, Legendre's pessimism was fully justified. But in the meantime, work in elasticity was developing rapidly in the hands of the most competent mathematicians of the day—Fourier, Navier, Poisson and Cauchy. As for Navier's work, Bucciarelli and Dworsky set the stage for growing conflict by noting that "Poisson had been upstaged in his own domain" (page 100). Navier, by making certain assumptions about forces between molecules, gave the first system of differential equations defining the behavior of a three-dimensional solid acted on by external forces. Poisson objected, criticizing Navier's assumptions about molecular forces as a step backwards into metaphysics. Poisson also disliked the use of Lagrange's variational techniques, which he felt were inappropriate in dealing with a subject that should be approached in a purely mechanical molecular way.

Navier's work had been submitted to the Academy for judgment (upon which publication would be approved), and Cauchy was assigned the paper. Instead of issuing an evaluation, however, Cauchy began to publish his own results on the subject, causing Navier to protest! But as Bucciarelli and Dworsky acknowledge, Cauchy's insights went much further than anyone else's at the time, and included salient definitions of stress, strain, derivations of equilibrium with respect to both, and to displacement of each point of an elastic body. In short, Cauchy managed to bring together all of the most essential features of today's theories of elasticity (page 102).

By now, the trend in elasticity was moving away from Germain on two fronts—Cauchy developing his ideas in terms of matter viewed as a continuum, Navier adopting the molecular approach of her rival Poisson (page 104). Germain nonetheless continued to study and publish on the subject, although her later efforts are also described by Bucciarelli and Dworsky as "slight," and "unsuccessful," (page 105). Work she submitted to the Académie was never formally evaluated because "her detailed analysis simply did not constitute first-rate work" (page 106); put more bluntly, the Academy simply could not approve work that was "so inadequate and trivial" (page 107).

In 1828 Poisson published a grand theory meant to treat elasticity in terms of a comprehensive molecular vision (page 109). Navier, in response, promptly criticized Poisson for failing to cite his work, and also noted Germain's "ingenious hypothesis" about the nature of elasticity. Poisson lost no time in pointing out the great confusion in Germain's work, and went on to criticize Navier's results, notably his boundary conditions. Navier, in rebuttal and in support of his friend, mentioned Lagrange's use of Germain's hypothesis to derive the plate equation. Even Germain herself felt obliged to respond, and in a general, philosophical commentary on the insufficient nature of the molecular hypothesis, she stressed her role in exposing its limitations. Published in the *Annales de Chimie* for 1828, her reply to Poisson struck a genuinely Newtonian note, typical of 19th-century positivism. She insisted that the object of mathematics was not the investigation of the causes of phenomena. One was restricted to the incontestable facts of elasticity. All she felt justified in assuming was the tendency of elastic bodies to return to their original shapes when stretched—nothing more need nor should be assumed about how molecules interact (page 110).

Towards the end of her life, Germain's circle of acquaintances apparently included Auguste Comte—father of Positivism—and even G. W. F. Hegel, when he visited Paris. But there seem to have been few contacts with any members of the Académie. Even so, she continued to write. Among her last scientific memoirs was one concerning the curvature of surfaces and a second on number theory. She wrote to Gauss about both, and the letter she sent sadly reflects the isolation of her life in Paris, where she saw few scientific colleagues and never enjoyed easy access to publications and journals. The letter, in fact, "was impressionistic and showed a lack of understanding of (Gauss') work" (page 116). It is little wonder that he did not reply.

Germain also found the energy to produce one last polemic on elasticity which was published in Crelle's *Journal* in 1830. In part, the uprising in Paris that year may have driven her back to

her studies, representing the same sort of refuge she had found in science during the early years of the French Revolution, when she was only thirteen. But now she was fifty-four, and there was no future for her; she was ill and complained of the discomfort and suffering she was tired of bearing. A year later, in 1831, she died of cancer.

To the extent that Germain was of special interest to her contemporaries, it was because she was a natural anomaly. Even Condorcet, the most liberal of the *philosophes* on the subject of women (he wrote a pamphlet *On the Admission of Women to the Sufferage*, anticipating John Stuart Mill's famous essay, *The Subjection of Women*, by nearly a century), wrote that women had shown no genius in the higher forms of science and philosophy, which required he said, extraordinary brain power. Germain showed signs of genius, scientific genius, and this made her unique.

For a time, especially at the beginning when many took Germain's early efforts in mathematics as signs of real promise, she enjoyed a substantial following, including the prominent Legendre, and later the young Italian mathematician Libri. Legendre, however, seems to have abandoned her as the consensus over her work on elasticity theory became increasingly negative, and Libri may well have been attracted to her because, in a sense, they were both foreigners and outsiders.

Libri, in fact, is an interesting case, for his relationship with Germain reveals something of the nature of intellectual life in Paris in the 1830's and raises some important questions about Germain's status there. Libri apparently knew Germain socially, at the end of her life. Both were interested in number theory, and Bucciarelli and Dworsky suggest that Germain became a kindly patron of sorts; eventually Libri went on to be her first biographer.

Libri saw Germain as a very social, urbane and noble conversationalist, but this may only have been his way of adulating his biographee. If she could be socially at ease, even adept, with the likes of Comte and Hegel, as he reports, why not on similar terms with the Parisian mathematicians, whose approval she repeatedly sought, and with whom she was willing to correspond whenever she wished clarification on theoretical points related to her work? Were they too cliquish, an exclusive group within the Academy that looked down on other Frenchmen doing important work, let alone others? Was Libri, as a foreigner himself, more at home with Germain than Legendre or others of the intellectual establishment interested in number theory?

What Sophie Germain needed, as her biographers agree, was an orderly education in mathematics and in judging the nature of mathematical questions. She needed criticism as well as encouragement, discipline as well as indulgence. Instead, she appears to have been regarded as a curiosity, a marvel. As Libri said of her, she was a "superior woman, who of all who pursued mathematical studies the farthest, the only one, to our knowledge, who has made real progress" (page 118). But in fact, where her work on elasticity was concerned, she lacked the mathematical facility to work with Lagrange's variational techniques, and as a result, she could not command the respect of the French mathematical intelligentsia of her day. Although faint praise was delivered, she could not hold her own place among the most powerful minds and members of the Académie. As Bucciarelli and Dworsky put it, "with no way of avoiding it, and through no fault of her own, she looks like a fool in the world of French science" (page 97).

As Bucciarelli and Dworsky stress, science then as now was rooted in human lives and the interplay between people and institutions. And yet, if anything, Germain's life illustrates the problems that resulted from her failure to interact successfully either with scientists or institutions. Although the Académie went so far as to confer a *prix extraordinaire* upon her, it did so reluctantly, and never opened its doors to her officially. Fourier, when he became Permanent Secretary of the Academy in 1822 (he was also a personal friend), arranged that she might attend any public meetings of the Institute that she wished to, with a reserved place in the center of the hall as a demonstration of "all the interest that your mathematical works inspire" (page 89). But Fourier's generosity in helping Germain also indicates the extent to which, formerly, she had been excluded from scientific meetings in Paris, despite the reputation she had acquired for herself.

But perhaps her greatest handicap was her self-imposed isolation, and her failure to have any systematic interaction with the scientific practitioners of her day, either directly, or through their writings. Had Lagrange been on hand, even periodically, to explain the complexities of double

integrals, to respond to her questions, to provide coaching, or discussion of esoteric matters related to the problem of elasticity, she might have been able to go much further. Unlike the Marquise de Chastelet or Madame de Staël, Sophie Germain seems to have worked alone—perhaps she even preferred to work alone. But this meant that she was deprived of suggestive or corrective interaction with her contemporaries. Moreover, she does not seem to have read carefully—or to have been able to find easy access to material that would have been of immense value. Why she was never able to profit from Lagrange's discussion of double integrals, presented in the new edition of his *Mécanique Analytique* in 1811, remains a mystery.

What Bucciarelli and Dworsky have written is a very interesting book on the subject of elasticity and its development during the early 19th Century. But Sophie Germain, as a person of flesh and blood, is virtually absent. As the authors admit, Germain literally "disappears" for lack of any detailed biographical information. The rest is rather meagre, and her "biographers" are left to survey her work, which for the mathematician is unfortunately disappointing. On the other hand, Germain's life, in the words of Bucciarelli and Dworsky, became her work, and this in turn "became a part of the working of the scientific community." It is here, on the level of French institutions and rivalries, rather than the biography of Sophie Germain, that their book succeeds best.

---

*Dynamical Systems on Surfaces*. By C. Godbillon. Translated from the French by H. G. Helfenstein. Springer-Verlag, Berlin-Heidelberg, Germany, 1983. 201 pp. $19.80.

R. F. Williams
*Department of Mathematics, Northwestern University, Evanston, IL 60201*

When I first heard this title, my mind traveled back over the last fifteen or so years of excitement in dynamical systems. It soon became clear exactly how much of this research has been just that: *dynamical systems on surfaces*. The recent resurgence of rational mappings, as dynamical systems, roughly centered around Dennis Sullivan, has three continents throbbing with the latest news about these self maps of the Riemann sphere. And not long ago, the "pseudo-Anosov maps" of Thurston broke upon us with a beautiful melding of dynamical systems and quasi-conformal mappings. They have proved to be an important tool in several fields.

Let us rebegin at the beginning. A dynamical system herein will be one of two types. First, a smooth autonomous ordinary differential equation (ODE) on a manifold usually without boundary. This means a *vector field* on a manifold. By the existence and uniqueness theorem, this is equivalent to a *flow* on the manifold. If one takes a section, i.e., a manifold of one less dimension, *transverse* to the flow, one obtains the *Poincaré map* of the flow which assigns to each point $x$ in the section, the "first return" $f(x)$ of the orbit through $x$ to the section. Thus the Poincaré map is a diffeomorphism $(f)$ defined on a manifold of one lower dimension. Iterating the Poincaré map corresponds exactly to taking the second-, third-, etc., returns of the flow to the section so that flow and map reflect the same properties to a large extent. Thus the second type of system is a *diffeomorphism*; evolution of an ODE corresponds to *iterating* a diffeomorphism. Abstractly, a flow is an action of the real line, $R$; a diffeomorphism generates an action of the integers, $Z$.

A system is *structurally stable* (1937, Andronov and Pontrjagin) provided that sufficiently small perturbations ($C^r$ topology) cause no structural change in the evolution of the system. A system which completely lacks stability would be a poor model for reality, as reality is *always* a perturbation of what we think it is. Thus some kind of stability is crucial.

The modern flowering of dynamical systems began when Peixoto proved that structurally stable flows are dense on surfaces, about 20 years ago. An earlier version (for the disk) helped stimulate Smale who quickly became a central force in dynamical systems. His work on "Morse-Smale" systems is largely motivated by examples on surfaces. But the "Smale horse shoe"

is truly 2-dimensional in conception.

This example is both beautiful and baffling; since I have seen many talented expositors try to give a simple explanation, I will give only an outline.

The basic notion is the *stable manifold*, and the simplest instance of this is a hyperbolic linear mapping of the plane; i.e., one with two real eigenvalues, one inside and one outside the unit circle. The *stable manifold* of such a map is the line in the eigen-direction corresponding to the smaller eigenvalue. It is characterized as the set of all points which under iteration of the map tend (exponentially) toward the origin. Dually, the *unstable* manifold consists of those points which tend away, under iteration and is the line in the other eigen-direction. All other orbits of points can be understood in terms of the unstable and stable manifolds, as they travel along hyperbolas, with these lines as asymptotes.

Now a *nonlinear* map leaving the origin fixed, and with derivative a hyperbolic map as mentioned above, will also have a stable and unstable manifold. At the origin these manifolds are tangent to the stable and unstable manifolds of the derivative, but far away they can bend depending upon the nonlinear map. Poincaré knew of examples of nonlinear maps in which the stable and unstable manifolds cross one another—and he knew that in such cases, all hell breaks loose. Such a point of intersection is called a *homoclinic* point of the fixed point; the orbit of a point of intersection is called a *homoclinic orbit*. Note that any homoclinic orbit tends to the fixed point under both forward and backward iterations of the map.

Poincaré felt that the resulting dynamics was so complicated as to defy analysis. But this is just what Smale did with an example on the 2-dimensional sphere now universally called the Smale "horseshoe" because it maps a square into a horseshoe shape, extending outside the original square. Smale proved that this example is structurally stable and codified its dynamical properties as being *isomorphic to the Bernoulli 2-shift*. This last is the probability space usually employed to describe the experiment of tossing a coin.

Smale's "horseshoe" marks the first step toward understanding "chaotic" or "strange" systems.

At about this time Smale wondered whether every dynamical system must have an "attractor." This word was introduced in about 1966 (Smale or Thom?) to mean a "generalized sink;" that is, an invariant set $A$ toward which all points in some neighborhood tend under iteration of the system. (An attractor should also have a dense orbit.) Thom pointed out that the linear hyperbolic map $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, when passed down to the torus, has no attractor. This map, so simple as to be linear, yet complicated enough to have a *global hyperbolic structure*, inspired Anosov's work on "Anosov diffeomorphisms," and later the pseudo-Anosov maps mentioned above.

In this case, a "global hyperbolic structure," is easily understood in terms of linear algebra. The dynamics are complicated because in essence, *all points are homoclinic*. Let $U$ denote the family of all lines parallel to the unstable eigen-direction in the plane, and let $S$ be those lines parallel to the stable eigen-direction. Both of these families foliate the plane and (as families) are invariant under the linear map; this map stretches the lines in $U$ and shrinks those in $S$. Now these families of lines pass down to families of lines on the torus because the eigen-directions have *irrational* slopes. Though each line is now densely winding throughout the whole torus, the lines are stretched or shrunk by the map on the torus just as before.

Though the appellation "strange attractor" came later, the first two "strange" attractors were discovered at the end of 1966 by Smale. The more famous Smale's "DA" or "derived from Anosov" map is a diffeomorphism of the torus with a 1-dimensional attractor. (It appears that until their discovery there had been no need for the concept "attractor.") Today the phrase "strange attractor" is widely used by physicists and engineers as well as by dynamicists. It was coined by Ruelle and Takens in the essentially infinite dimensional context of turbulent flow. But much of the current excitement concerns—you guessed it—a two dimensional system, the "Hénon attractor." This is a diffeomorphism of the plane in which the two coordinate functions are polynomials in $x$ and $y$, one linear, the other quadratic.

When I finally had Godbillon's book in my hands, I was surprised to find that *none* of these

seven topics was mentioned. In all fairness, this is partly because Godbillon's book is elementary. But the *real* reason is that this is a book about *foliations*, not dynamical systems. True, titles often reflect fashion more than content. But I cannot help regretting the demise of the book this title conjured up in my mind.

*Dynamical Systems on Surfaces* is in Springer's "Universitext" series. And this book could be used as a text leading up to more advanced material on foliations.* The prospective teacher is warned that there is no index and that the exercises are spotty, both in frequency and to a lesser extent, relevance.

The most glaring omission is the stable manifold theorem (see p. 104). In this connection, Godbillon apologizes that there is a missing chapter—"which would have been dealing with structural stability." But there is no way to do structural stability without the stable manifold theorem, and to include this missing chapter would have meant rewriting the whole book.

In summary, this is an elementary introduction to such topics as flow boxes, rotation numbers, stability criterion near singular points, the index of a vector field, and geometric analysis of singular points. Hartman's theorem, that a smooth hyperbolic map is locally conjugate to its linear part near its fixed point is proved but only in the contracting case. Attracting periodic orbits are discussed, but not the "in phase" theorem.

The "in phase" theorem is less widely understood than it should be; it is a beautiful result. A periodic orbit $A$ of an ODE is said to be *attracting* provided it has a neighborhood within which all orbits tend toward $A$ as required by our general definition of attractor above. Taking a Poincaré section yields a diffeomorphism with fixed point where the section intersects the periodic orbit $A$. If the Poincaré map is a contraction (this is independent of the choice of section) at the fixed point, then $A$ is an attracting periodic orbit. Godbillon goes this far.

But the same hypothesis on the Poincaré map implies much more: each point in this neighborhood is "in phase" with some point on the periodic orbit, so that not only do nearby orbits tend to look periodic after a while, but in addition, each point in a neighborhood is associated with or "in phase" with a *certain* point on the periodic orbit. The set of all points "in phase" with a point $x$ is the (strong) stable manifold of $x$, and has codimension one. These strong stable manifolds *foliate* the neighborhood, and the flow moves leaves of the foliation along into other leaves.

The Denjoy theorem about $C^2$ or better flows on the torus and his counter example for $C^1$, is given, partly as a sequence of exercises. But along with these there is much technical material designed for the specialist in foliation theory. And much, much more, central to dynamical systems, is left out.

*This reviewer recommends Palis-deMelo, *Geometric Theory of Dynamical Systems*, Springer, 1982, for a course in dynamical systems.

---

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

Professor Robert C. Thompson's essay on inflation [5] does not reflect the mathematical literature concerning his subject, specifically numerous undergraduate texts, for instance Wonnacott and Wonnacott [8, p. 274]. Professor Thompson's "true growth rate" is in fact the

Fisher effect, published in 1896 by the famous economist and mathematician Irving Fisher [**1**], and still read by economics students in his classic reference masterpiece [**2**, p. 39]. Indeed, with $r$, $i$, and $g$ denoting the rates of inflation, interest, and growth of purchasing power—Professor Thompson's "true growth rate"—the Fisher effect states precisely that $g = (i - r)/(1 + r)$.

Professor Thompson also asserts in [**5**, p. 208] that this formula "isn't known to a good many professionals in finance [...]." Evidence to the contrary abounds: introductory economics and finance classes learn the Fisher effect from texts such as Wonnacott [**7**, p. 298] and Van Horne [**6**, p. 125], respectively. Even engineers see this economic model, for example in Newnan's textbook [**4**, p. 334]. Both Newnan and Wonnacott present numerical examples, and one would have illustrated Professor Thompson's composition quite nicely for students. Actuaries, too, derive the relation $g = (i - r)/(1 + r)$ when they study their classic text by Kellison [**3**, Problem 13, pp. 26 & 235] to drill for the fourth of their ten professional examinations. Then they apply Fisher's formula to *annuities-immediate in geometric progression* [**3**, p. 85, Equation (4.40)], which Professor Thompson calls "strings of inflating payments" [**5**, p. 209].

**References**

**1.** Irving Fisher, Appreciation and interest, Publications of the American Economic Association, Third Series, vol. XI, no. 4 (August 1896), pp. 331–442.

**2.** ______, The Theory of Interest, Macmillan, New York, 1930.

**3.** Stephen G. Kellison, The Theory of Interest, Irwin, Homewood, IL, 1970.

**4.** Donald G. Newnan, Engineering Economic Analysis, 2nd ed., Engineering Press, San Jose, CA, 1983 (1st ed. 1976).

**5.** Robert C. Thompson, The true growth rate and the inflation balancing principle, this MONTHLY, vol. 90, no. 3 (March 1983), pp. 207–210.

**6.** James C. Van Horne, Financial Market Rates and Flows, Prentice-Hall, Englewood Cliffs, NJ, 1978.

**7.** Paul Wonnacott, Macroeconomics, Irwin, Homewood, IL, 1974.

**8.** Paul Wonnacott and Ronald Wonnacott, Economics, 2nd ed., McGraw-Hill, NY, 1982.

Yves Nievergelt
Department of Mathematics
University of Washington
Seattle, WA 98195

## ACKNOWLEDGMENT

The editors thank the following persons who have refereed manuscripts between mid-1983 and mid-1984:

Harvey L. Abbott, Janos Aczel, C. D. Aliprantis, Richard J. Allen, Donald E. Amos, Glen Anderson, Geogre E. Andrews, James R. Angelos, Tom Apostol, Judith Arms, Barry C. Arnold, Kenneth Arrow, Richard A. Askey, Raymond G. Ayoub, E. A. Azoff, John M. Bachar, B. J. Ball, David W. Barnetti, Robert G. Bartle, Paul T. Bateman, Martin Bendersky, John J. Benedetto, Max Benson, Russell V. Benson, S. K. Berberian, Peter G. Bergmann, Bruce C. Berndt, Frank R. Bernhart, David Blair, W. W. Bledsoe, Jerry L. Bona, Richard C. Bradley, S. J. Brams, David M. Bressoud, John D. Brillhart, Duane M. Broline, Arlen Brown, Cynthia A. Brown, Ezra Brown, Richard Brown, Robert F. Brown, Richard A. Brualdi, Andrew M. Bruckner, Aiden A. Bruen, Edwin Buchman, Owen Burkinshaw, Paul F. Byrd, Gerald T. Cargo, Phyllis Joan Cassidy, Frank S. Cater, Judith N. Cederberg, Gulbank D. Chakerian, Orin N. Chein, F. R. K. Chung, Marshall M. Cohen, Paul M. Cohn, Courtney Coleman, Charles C. Conley, P. C. Consul, John B. Conway, Kenneth L. Cooke, Clifton Corzatt, Carl Cowen, H. S. Macdonald Coxeter, Henry H. Crapo, Frederick H. Croom, James J. Cross, Joseph Crosswhite, Morton L. Curtis, Thomas W. Cusick,

Jiri Dadok, Ulrich Daepp, Chandler Davis, Martin D. Davis, Jane M. Day, Stanley R. Deans, Herbert A. DeKleine, Charles R. DePrima, Leroy J. Dickey, J. Diestel, Seymour Z. Ditor, Nathan J. Divinsky, John D. Dixon, Jack J. Dongarra, Robert S. Doran, Vladimir Drobot, Lester E. Dubins, Underwood Dudley, Gerald A. Edgar, Charles C. Edmunds, Harold M. Edwards, Stewart Either, Ernest Enns, Paul Erdös, R. J. Evans, K. J. Falconer, William Firey, Evan Fisher, Harley Flanders, William Fleissner, Peter Fletcher, Denis R. Floyd, Lester R. Ford, Jr., David A. Foulser, George K. Francis, Joel N. Franklin, Michael L. Fredman, John A. Fridy, Joseph Gallian, Steven Galovich, Fred Galvin, Carl Ganser, Martin Gardner, George Gasper, Jr., Bernard R. Gelbaum, Ira M. Gessel, John S. Giles, William B. Giles, Michael J. Gilpin, Samuel Goldberg, Herbert Goldstein, Larry J. Goldstein, Solomon W. Golomb, J. E. Goodman, Sandy Grabiner, Ronald L. Graham, Jack E. Graver, Marvin Greenberg, Branko Grünbaum, H. W. Guggenheimer, Victor W. Guillemin, Ranee Gupta, Darrell E. Haile, Thomas G. Hallam, Leon Harkleroad, Hiroshi Haruki, William S. Hatcher, T. W. Hawkins, Louise Hay, Garry Helzer, Leon A. Henkin, Michael Henle, James V. Herod, Reuben Hersh, I. N. Herstein, Herbert W. Hethcote, Edwin Hewitt, Richard Hill, Norman Hills, Peter G. Hinman, Shu-Ping Hodgson, David C. Hoff, Richard B. Holmes, Ross A. Honsberger, Frank C. Hoppensteadt, Ralph Howard, Roger C. Howe, Karel Hrbacek, James E. Humphreys, Dean Isaacson, John R. Isbell, D. M. Jackson, Lloyd K. Jackson, James E. Jamison, Gerald J. Janusz, David L. Johnson, Dennis Johnson, Barbara Kaiser, Eva Kallin, Nigel Kalton, David W. Kammler, Edwin H. Kaufman, Nicholas D. Kazarinoff, H. Jerome Keisler, Kenneth R. Kellum, Paul Kelly, John O. Kiltinen, Clark H. Kimberling, Henry C. King, John R. Kinney, S. C. Kleene, Daniel J. Kleitman, Morris Kline, Martin Kneser, Arthur R. Knoebel, Donald E. Knuth, Neal I. Koblitz, Joseph D. E. Konhauser, Nicholas Jacob Korevaar, Czes Kosniowski, Erwin O. Kreyszig, Richard P. Kubelka, David W. Kueker, Thommas J. Laffey, Jeff Lagarias, Lester H. Lange, Loren C. Larson, Terry D. Lenker, Frank D. Lesley, Mark D. Levin, John S. Lew, Andy Liu, Peter A. Loeb, David E. Logothetti, Warren Loud, Thomas R. Lucas, Erwin Lutwak, W. A. J. Luxemburg, Roger C. Lyndon, Saunders Mac Lane, Michael Makkai, J. A. Marcus, Mario Umberto Martelli, John Masterson, Kevin McCurley, Gerard McDonald, Jack E. McLaughlin, D. G. Mead, H. Mehrtens, Thomas J. Miles, Martin H. Millman, Susan Montgomery, David S. Moore, Daniel Moran, James S. Muldowney, Jan Mycielski, John Myhill, Isaac Namioka, Vladomir A. Naroditsky, C. S. J. A. Nash-Williams, Ivan Niven, Andrew M. Odlyzko, John M. H. Olmsted, Barbara L. Osofsky, John C. Oxtoby, Edgar Palmer, Theodore W. Palmer, George D. Parker, Emanuel Parzen, E. M. Patterson, Das Peddada, Jean J. Pedersen, Gyorgy Petruska, John A. Pfaltzgraff, Richard E. Pfiefer, Eswar G. Phadia, Richard Phillips, Veril L. Phillips, George Piranian, Jacob M. Plotkin, William F. Pohl, Harry Pollard, Carl Pomerance, Sidney C. Port, Mohsen Pourahmadi, Stephen E. Puckette, Hilary Putnam, Joseph E. Quinn, M. Rabin, Heydar Radjavi, Juerg J. Raetz, David F. Rearick, Howard E. Reinhardt, Clive M. Reis, Bruce A. Reznick, J. Ian Richards, Fred Richman, A. Wayne Roberts, Nick J. Rose, Marvin Rosenblum, George Rosenstein, Kenneth A. Ross, Peter Ross, J. J. Rotman, Kenneth R. Rotman, James L. Rovnyak, Halsey L. Royden, Lee A. Rubel, Walter Rudin, D. G. Saari, Cora Sadosky, G. W. Sands, Jonathan Schaer, J. J. Schäffer, Martin G. Scharlemann, Bruce L. R. Shawyer, Glenn E. Schober, John Schue, Allen Schweinsberg, Charles Seebeck, Jeffrey O. Shallit, Lawrence A. Shepp, Gary J. Sherman, Neelima Shrikande, D. A. Singer, John H. Skillings, William Sledd, Daniel C. Sloughter, David Smith, Douglas D. Smith, H. Paul Smith, Martha Smith, Craig Smoryński, Kenneth W. Spackman, Jerome Spanier, Ivar Stakgold, D. W. Stanton, Norton Starr, Nick Stavrakas, Andrew Sterrett, Michael Stob, Kenneth B. Stolarsky, Gilbert Strang, Volker Strassen, Robert S. Strichartz, Karl R. Stromberg, Mathukumalli Ventana Subbarao, David B. Surowski, Howard S. G. Swan, Lajos F. Takacs, Robert E. Tarjan, W. T. Taussky, S. J. Taylor, Robert C. Thompson, Dmitri E. Thoro, John Todd, Thomas Tucker, W. T. Tutte, Charles L. Vanden Eynden, Eric van Douwen, Samuel S. Wagstaff, Jr., Jeanne Kerr Wald, Edward T. H. Wang, Clifford Weil, Howard J. Weiner, J. G. Wendel, Gregory P. Wene, Albert Wilansky, J. B. Wilker, Hugh C. Williams, S. Winograd, John Wolfskill, Daniel Wulbert, Lawrence A. Zalcman, Frank A. Zorzitto.

HM

## Basic Mathematics and Algebra with Applications Fourth Edition

**M. Wiles Keller,** Purdue University
**James H. Zant,** Oklahoma State University
546 perforated pages • paper
Instructor's Annotated Edition
Alternate Testing Program • 1984

Keller and Zant's developmental mathematics program reviews both basic mathematics and elementary algebra within a single text. The program helps students help themselves in three important ways:

- *Step-by-step instruction.* Step-by-step, annotated worked examples show the student how to work algorithms and solve problems.
- *Objective-referenced organization.* The text is organized around a hierarchy of 141 learning objectives, which link lessons, practice sets, and test items for easy management of instruction.
- *Expanded applications-oriented content.* An abundance of problems in situational settings provides the student with an opportunity to use skills as they are learned.

## Handbook of Basic Math with Applications

**James H. Zant** and **M. Wiles Keller**
224 perforated pages • paper • Answer Booklet • 1984

An economical review of the basic math portion of the above text.

## Essential Mathematics with Applications

**Vernon C. Barker** and **Richard N. Aufmann**
Both of Palomar College
286 pages • paper • Answer Booklet • 1982

## Programming for Microcomputers: Apple II BASIC

**June Grant Shane,** Indiana University
432 pages • spiralbound • Grid Sheet Booklet • 1983

## PCSP Personalized Computational Skills Program Brush-Up Books®

**Bryce R. Shaw,** Senior Author
**Gwen Shelton,** Bee County College, Texas
**Sandra Pryor Clarkson**
Hunter College, New York
Contributors: **Richard N. Aufmann**
and **Vernon C. Barker**
Both of Palomar College
**Linda M. Oldaker,** Arizona State University
Adviser: **Ernest J. Cioffi**
University of Southern California

**VOLUME 1 COMPUTATIONAL SKILLS**
**Module A: Whole Numbers**
**Module B: Fractions**
**Module C: Decimals, Proportions, Percents**
**Complete Three-Module Text**
544 perforated pages • paper • Instructor's Manual including Alternate Testing Program • 1980

**VOLUME 2 APPLICATIONS**
**Module D: Applications Using Computational Skills**
**Module E: Special Skills with Applications**
**Module F: Beginning Algebra with Applications**
**Complete Three-Module Text**
512 perforated pages • paper • Instructor's Manual including Alternate Testing Program • 1981

**COMBINED EDITION**
**SKILLS AND APPLICATIONS**
480 perforated pages • paper • Instructor's Manual including Alternate Testing Program • 1982

THE AMERICAN
MATHEMATICAL MONTHLY

olume 92, Number 2 February 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, University of California, Berkeley, CA 94720. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# AUTOMATED REASONING*

L. WOS

*Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439*

**1. Introduction.** Two of the principal activities of mathematicians and logicians are the search for a proof of some lemma or theorem and the search for some model or counterexample. Until recently, both activities were conducted by the researcher alone or conducted with a colleague or colleagues with a similar interest. Recent occurrences suggest that another alternative now exists. The alternative is that of relying on the assistance of a computer program, an automated reasoning program. One such automated reasoning program, AURA (for AUtomated Reasoning Assistant) [**12**], has in fact provided valuable assistance in answering certain previously open questions from the fields of ternary Boolean algebra [**14**], finite semigroups [**16**], [**17**], and equivalential calculus [**15**], [**21**]. The capacity of this program both to find proofs and to generate models and counterexamples played a vital role in answering the open questions. AURA is not the only automated reasoning program that has been used to obtain new results in mathematics. Slightly more than 15 years ago, a reasoning program [**3**] of a distinctly different design proved a lemma in lattice theory that had not been previously known. Just as the computer has been used for years to assist in certain numerical aspects of mathematics, it can now be used to assist in various reasoning aspects.

Since an automated reasoning program can provide such valuable assistance in diverse fields of mathematics and formal logic, you might wish to know more about it.

1. What is automated reasoning?
2. What has been achieved with the assistance of an automated reasoning program?
3. How much work is required to use such a program?
4. Is such a program available?
5. Which types of research are amenable to assistance from such a program?

By answering these questions, I intend to satisfy the curiosity of some, pique the curiosity of others, and, most important, enlist the aid of both mathematicians and logicians in finding additional open questions to attack with the assistance of an automated reasoning program. Equally, I wish to interest members from various fields to begin actively using a reasoning program as a research assistant.

**2. The Elements of Automated Reasoning.** Automated reasoning is concerned with the discovery, formulation, and implementation of concepts and procedures that permit the computer to be used as a reasoning assistant. One of the primary objectives is the design and implementation of a computer program—often called a theorem-proving program—that automates the process known as reasoning. By reasoning, we mean the process of drawing conclusions rigorously, as in mathematics and in formal logic. Reasoning here means logical reasoning, not probabilistic or common-sense reasoning. When a conclusion is drawn from some given set of assumptions, the

---

Lawrence Wos received his Ph.D. in mathematics under Reinhold Baer at the University of Illinois in 1957. Since then, he has conducted research in various aspects of automated theorem proving and automated reasoning at Argonne National Laboratory, Argonne, Illinois 60439. He introduced the notion of "strategy" into the field. In collaboration with one of his colleagues, he answered (previously) open questions in abstract algebra and in formal logic. The questions were answered by relying heavily on the automated reasoning program AURA developed jointly at Argonne and Northern Illinois University. In recognition of his achievements, Wos (together with his colleague) was awarded the Mathematics Prize for the best current research in automated theorem proving.

*This work was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the U.S. Department of Energy under Contract W-31-109-ENG-38.

conclusion must follow logically and inevitably. If the conclusion itself represents false information, then one of the assumptions from which it was obtained is false. How does an automated reasoning program derive its logical power and rigor?

A program of the type under discussion reasons by applying well-defined inference rules to statements called "clauses". (The discussion of automated reasoning programs given here is influenced by our particular programs. Certain features, for example the use of "clauses" and of "weighting", are often not available, especially in those programs with a quite different design.) For the concept of a "clause", first note that, informally, an "atom" is a statement free of logical operators. A "literal" is an atom or an atom preceded by the "not" operator. Finally, a "clause" is the "or" of a (possibly empty) finite set of literals. For example, the axiom of left identity in a group—the product of $e$ and $x$ is $x$ for all $x$—can be written as the clause

$$\text{PRODUCT}(e, x, x),$$

which contains a single literal. Where "$-$" denotes "not" and "$|$" denotes "or", the statement "the subset $H$ is closed under addition" can be written as the clause

$$-H(x) \mid -H(y) \mid H(\text{sum}(x, y))$$

in which the first two literals are called "negative literals". This clause, which can be read "if $x$ and $y$ are in $H$, then their sum is in $H$", is obtained by using the fact that "if $P$, then $Q$" is logically equivalent to "not $P$ or $Q$".

Each inference rule selects from the available information some number of statements, each of which satisfies certain syntactic requirements, and yields a conclusion only if it satisfies certain (possibly different) syntactic requirements. For example, one inference rule employed by reasoning programs yields a conclusion only if it contains a single literal. Another inference rule requires that the conclusion contain only positive literals, placing no restriction on their number. Many inference rules that are employed are generalizations of the familiar rules of modus ponens and syllogism. Still others generalize equality substitution. For example, an inference rule exists that, when applied to the equations "$x + -x = 0$" and "$y + (-y + z) = z$", yields in a single step "$y + 0 = -(-y)$". In clause form, from

$$\text{EQUAL}(\text{sum}(x, \text{minus}(x)), 0)$$

and

$$\text{EQUAL}(\text{sum}(y, \text{sum}(\text{minus}(y), z)), z)$$

the clause

$$\text{EQUAL}(\text{sum}(y, 0), \text{minus}(\text{minus}(y)))$$

is obtained. To be acceptable, an inference rule must be sound: the conclusions obtained with it must be logical consequences of (the conjunction of) the statements to which it is applied to yield them. From the viewpoint of model theory, an inference rule is sound if, when it is applied to a set of premisses to yield a conclusion, the conclusion holds in any model that satisfies the premisses from which it is obtained.

When an automated reasoning program draws a conclusion, as in the spirit of high school geometry, it cites the specific and immediate ancestors of the conclusion, thus enabling you to check its work. All steps of a chain of reasoning are made explicit; you can check any or all steps of a deduction simply by executing the same algorithm that was executed by the reasoning program. Any proof obtained by a reasoning program can, therefore, be refereed in a straightforward though tedious manner.

But how does a reasoning program recognize that a proof has been obtained? By far the most common form of proof obtained by an automated reasoning program is a proof by contradiction. As often occurs in mathematics, you begin by stating the theorem under study and assume it to be false. If the symbolic form of the theorem is "if $P$, then $Q$", you assume $P$ true and $Q$ false.

Should "if $P$, then $Q$" be a theorem, making this assumption yields a self-contradictory set of statements. Such a set of statements is called an "unsatisfiable" set. In the typical case, an automated reasoning program is presented with an unsatisfiable set of statements and asked to establish the unsatisfiability. Such a program recognizes it has completed a proof when it finds two statements that obviously contradict each other. For example, the program might eventually deduce that, for some element $a$, $a$ is not equal to $a$. Since the reflexivity axiom of equals, $x = x$, is ordinarily included in the submission of a problem to a reasoning program, deducing that $a$ is not equal to $a$ immediately signals a contradiction.

How does a reasoning program search for a proof, given statements that correspond to assuming $P$ true and $Q$ false? After all, in mathematics and in logic a random and undirected search seldom succeeds. Since the computer is so fast, does a reasoning program simply conduct an exhaustive search until the contradiction is discovered, discovered more or less by accident? By no means, for an attempt to prove even the most uninteresting theorems with an exhaustive search produces a tremendous number of facts, almost all of which are irrelevant. Even though the computer is very fast, such a search requires far too much time. What occurs instead, when using an automated reasoning program, is the employment of "strategy". Some strategies, called "ordering strategies", direct a reasoning program's attack on a problem, employing criteria for selecting the information on which to focus next. Some strategies, called "restriction strategies", restrict a reasoning program's attack, totally prohibiting certain paths from being explored.

An example of an ordering strategy is "weighting" [**9**]. You can assign "weights" to the various concepts in a problem, enabling a reasoning program to give priorities to the concepts. The program selects the information on which to focus next according to these priorities, selecting the least complex information for focus. You can assign weights to define "complexity" strictly in terms of the number of symbols in an expression. On the other hand, you can choose the opposite strategy—define "complexity" so that the more symbols in an expression, the more it is preferred for being considered next. As another alternative, you can assign weights so that multiplication is preferred over sum, so that inverse is preferred over identity, and so that the cube of an element is preferred over its square. Weighting allows you to impose your knowledge and intuition on the program's attack on a problem.

An example of a restriction strategy is the "set of support strategy" [**18**]. This strategy is in effect often employed by mathematicians and logicians, although not with an official name. The strategy attempts to take full advantage of the assumed consistency or satisfiability of the set of statements that corresponds to $P$, where the theorem under consideration has the form "if $P$, then $Q$". That set of statements—call it $P'$— is satisfiable, for its correspondent $P$ typically consists of the axioms for some domain of inquiry together with certain additional hypotheses. For example, if you were studying Boolean rings—rings in which, for all $x$, the product of $x$ with itself is $x$—then the set $P$ would consist of the axioms for a ring together with a statement that the ring is Boolean.

Imagine that you wish to prove that such a ring is commutative, that $xy = yx$ for all $x$ and $y$. You would begin by assuming the theorem false—assuming that there exist elements $a$ and $b$ in the ring with $ab$ not equal to $ba$. You would then have three sets of clauses—$A'$ which consists of the clauses that correspond to the axioms for a ring, $S'$ which consists of those that correspond to the Boolean property, and $D'$ which consists of those that correspond to assuming such a ring is not commutative. The set $P'$ would then be the union of $A'$ and $S'$.

The idea behind the set of support strategy is to prevent an automated reasoning program from randomly selecting clauses from $P'$ from which to draw conclusions. Such a random search for new information simply explores the theory of Boolean rings, rather than concentrating on the theorem to be proved, namely, such rings are commutative. Further, since the objective is to obtain a proof by contradiction, and since this forward search from $P'$ produces additional consistent information, the proof will be found only by accident— by accidentally proving commutativity. When used as recommended, the set of support strategy restricts the actions of a

reasoning program by preventing it from applying an inference rule to a set of premisses all of which are in $A'$, the correspondent of the set of axioms.

To use the set of support strategy, you choose a subset $T$ of the set $S$ of clauses that represents the problem to be studied. In the spirit of the example just discussed, the recommended choices for $T$ are either the union of $S'$ and $D'$, or $D'$ alone. For a set of premisses from among the elements of $S$ to be considered by an inference rule, the set is required to contain at least one element of $T$. By imposing the strategy on a reasoning program's search for information, the program is forced to key (at least recursively) on the elements in the chosen subset $T$. In the example from Boolean rings, the recommended choices for $T$ cause the program (recursively) to key on the Boolean property or on the fact that $ab$ is assumed different from $ba$. All new information is therefore traceable to $S'$ or to $D'$. If any lemmas from ring theory are supplied as part of the problem description, in keeping with the intent of the strategy, they are not included in $T$. The choice of $T$ as the union of $S'$ and $D'$ is reminiscent of the practice in mathematics and logic, for it causes an automated reasoning program to emphasize the role of the special hypothesis of the theorem—to key on those statements that are part of $P$ but are not the axioms and lemmas that are supplied.

An automated reasoning program of the type under discussion relies on two additional processes to enhance its effectiveness in finding proofs and in generating models and counterexamples. The first process, known as "subsumption" [**10**], enables a reasoning program to prefer general information to specific information. When presented with two statements such that one is captured by the other as a trivial corollary, the program discards the corollary. For example, in the ring problem cited earlier, if the program discovers that $x + x = x$ (for all $x$) and that $a + a = a$, then the second discovery is discarded in favor of the first. (Typically, the lemma that $x + x = x$ for all $x$ is in fact found and used in the proof that Boolean rings are commutative.) The order of the discoveries is not relevant, only the relative generality of the two statements. By means of subsumption, an automated reasoning program performs as a mathematician or logician often does, retaining the theorem and discarding the trivial corollaries.

The other process that enhances effectiveness is "demodulation" [**20**]. Through demodulation, a reasoning program normalizes and simplifies information. For example, with the appropriate information present, the program will replace $0 + x$ by $x$, $-(-x)$ by $x$, $(xy)z$ by $x\,(yz)$, and $xy + xz$ by $x(y + z)$. In other words, demodulation enables a reasoning program, if so instructed, to remove the (additive) identity from expressions, to cancel the inverse of the inverse, to right associate all expressions, and to use distributivity to write certain expressions in a factored form. The information for normalizing and simplifying can be supplied to the program at the beginning, and this information can be added to or found by the program as it conducts its search for the proof or for the model or counterexample. Despite the existence of a number of as yet unanswered questions concerning the use of demodulation, the process is a most valuable and often necessary feature for solving problems.

**3. Successes.** The attempt to answer open questions with the assistance of an automated reasoning program often benefits both fields—the field of mathematics or logic from which the problem is selected, and the field of automated reasoning. The various obstacles that are encountered, if overcome, lead to enhancements of existing techniques or even to totally new techniques for attacking problems. The value of finding open questions that are amenable to consideration by a reasoning program cannot be overestimated. In fact, your assistance in the search for such open questions is strongly encouraged. Equally, your use of a reasoning program in your own research could be beneficial to both mathematics and to automated reasoning itself.

The first open question that was answered with the aid of the automated reasoning program AURA (AUtomated Reasoning Assistant) [**12**] concerns the possible dependence of certain axioms for a ternary Boolean algebra [**14**]. A ternary Boolean algebra is a nonempty set satisfying the following five axioms, where the function $f$ can be thought of as a three-place "product" and $g$ as inverse:

(1) $f(f(v,w,x),y,f(v,w,z)) = f(v,w,f(x,y,z))$
(2) $f(y,x,x) = x$
(3) $f(x,y,g(y)) = x$
(4) $f(x,x,y) = x$
(5) $f(g(y),y,x) = x$

The precise question to be answered is: Are any of the first three axioms independent of the remaining four? (The fourth and fifth were already known to be dependent.) Each of the first three was proved independent by finding a model satisfying the remaining four but failing to satisfy the axiom in question. It was this study of ternary Boolean algebra that led to the technique for generating models and counterexamples [**13**] using an automated reasoning program.

The next open question that was answered [**16**], [**17**] with the assistance of AURA concerns the possible existence of a finite semigroup admitting certain mappings in combination. Specifically, does there exist a finite semigroup admitting a non-trivial antiautomorphism but admitting no non-trivial involutions? An antiautomorphism is a one-to-one onto mapping $h$ with $h(xy) = h(y)h(x)$. An involution is an antiautomorphism whose square is the identity. To be non-trivial, the mapping must not be the identity map. Commutative semigroups are of course not of interest for, in such a semigroup, an antiautomorphism is simply an automorphism. There does exist a non-commutative semigroup with the desired properties, and the first found has order 83 [**16**]. The smallest such semigroup [**17**] has order 7, and there are four such semigroups. Since there are over 800,000 semigroups of order 7, finding those four could present serious difficulties were one to simply sort through them looking for appropriate semigroups. The open questions of existence and minimality were answered by using generators and relations, relying on multiplication tables, considering possible isomorphisms, and applying "without loss of generality" arguments. The capacity of AURA both to find proofs and to generate models was required.

Other questions that were answered [**15**], [**21**] by relying on AURA concern the classification of formulas in the equivalential calculus. The equivalential calculus is that field of formal logic that studies the abstraction of "equivalence". In particular, the formulas

(1) $E(x,x)$ (reflexivity)
(2) $E(E(x,y), E(y,x))$ (symmetry)
(3) $E(E(x,y), E(E(y,z), E(x,z)))$ (transitivity)

taken together can serve as a complete set of axioms for the calculus. (The designations of reflexivity, symmetry, and transitivity, not ordinarily used in equivalential calculus, are chosen to provide an intuitive understanding.) The elements of equivalential calculus are the expressions that can be recursively well-formed from the variables, $x, y, z, w, \dots,$ and the two-place function $E$. In the equivalential calculus, there exist individual formulas that are sufficiently strong that they serve as single axioms for the entire calculus. The shortest single axiom for the calculus contains 11 symbols, excluding grouping symbols and commas. Of the 630 such formulas of length 11, at the time we began our study, eleven were known to be shortest single axioms and seven remained unclassified with respect to this property. We decided to consider the corresponding open questions: Are any of the following seven formulas strong enough to serve as a single axiom for the equivalential calculus?

(1) XJL $= E(x,E(y,E(E(E(z,y),x),z)))$
(2) XKE $= E(x,E(y,E(E(x,E(z,y)),z)))$
(3) XAK $= E(x,E(E(E(E(y,z)),x),z),y))$
(4) BX0 $= E(E(E(E(x,E(y,z)),z),y),x)$
(5) XCB $= E(x,E(E(E(x,y),E(z,y)),z))$
(6) XHK $= E(x,E(E(y,z),E(E(x,z),y)))$
(7) XHN $= E(x,E(E(y,z),E(E(z,x),y)))$

In particular, it was conjectured that no additional shortest single axioms existed—that the remaining unclassified seven formulas are each too weak to serve as a single axiom. The study revealed that the first five are each too weak, but, contrary to the conjecture, each of the last two is in fact strong enough [**15**] to be a single axiom. To obtain these results, the model generation technique used successfully to answer the preceding questions was deemed unsuitable. The approach we chose was to seek a complete characterization of all deducible theorems from each formula under study. The characterizations were obtained by devising a means for an automated reasoning program to employ schemata [**21**]. Again the study of an open question led to a new technique for using an automated reasoning program. The proof that XHK is a single axiom consists of 87 steps, including steps containing 71 symbols not counting grouping symbols and commas. The proof that XHN is a single axiom consists of 162 steps, including steps containing 103 symbols. An indication of the difficulty of completing these proofs is provided by the fact that there are more than 100,000,000,000 formulas containing 27 symbols.

The program AURA, as well as various other automated reasoning programs, has also been used in various studies only distantly related to mathematics and formal logic. The program has been used to design logic circuits, some superior to those already known. It has been used to validate the design of existing circuits also, and to experiment with the operation of control systems. Reasoning programs have been used to prove given claims for computer programs. Finally, the possible use of such a program for chemical synthesis is currently being investigated.

**4. Use and Availability.** The biggest obstacle to using an automated reasoning program is language. As with various fields of mathematics and of logic, automated reasoning has its own notation. The most common notation relies on the use of "clauses" [**19**]. The clause language is very closely related to the first-order predicate calculus. The language treats variables as meaning "for all", so it is trivial to say that "the square of every element is the identity", for example. When you wish to express that something "exists", depending on the nature of the statement, you employ a function of the appropriate number of variables or employ a constant. In addition, the conjunctive normal form is used—all statements are implicitly assumed to be separated by "and", while all literals of a statement are separated by "or". The only other operator that is acceptable is "not". Nevertheless, the language is more than sufficient for attacking a wide variety of problems. In fact, the language of clauses is extremely powerful, as we discovered in the past few years. Relationships involving "if then" and "equivalent" present no problem, for they are simply transformed into statements relying on "or", "not", and "and" having the same meaning. Despite the power and richness of the language, the presentation of a problem to an automated reasoning program is an art.

In addition to the representation problem, there are closely related problems of choosing the appropriate rules of reasoning and of choosing the strategies to control the rules [**19**]. Good choices of representation, inference rule, and strategy are extremely interconnected. The choices must not be made separately. Then there is the problem of providing the appropriate normalization and simplification information. Even with these obstacles, as evidenced earlier, an automated reasoning program can be successfully used to solve problems, some of which are quite difficult. The value of such an automated assistant can be seen by noting that the researchers who solved the open questions knew remarkably little about the areas from which the problems were selected.

You need not be an expert in automated reasoning to use such a program. The studies in circuit design, circuit validation, operation of control systems, and chemical synthesis are being conducted by people with little knowledge of automated reasoning. The people in question were able to initiate these studies by relying on manuals and by consulting with experts in automated reasoning. Even with the various successes and the current activities, we must again say that the use of an automated reasoning program is at present an art.

Until recently, the available written material consisted of manuals and formal texts on the

subject. The book "Symbolic Logic and Mechanical Theorem Proving" by Chang and Lee [**2**] is a very readable and rather formal introductory text. It covers well the topics of representation, inference rule, strategy, demodulation, and subsumption. The book "Logic for Problem Solving" by Kowalski [**4**] is somewhat less formal, and is particularly valuable for questions of representation. It covers applications such as state-space problems, and introduces the topic of logic programming with specific attention to the programming language of Prolog. The book "Automated Theorem Proving: A Logical Basis" by Loveland [**5**] presents a complete and thorough treatment and in a very formal manner. It contains much material on the logical foundations of the subject, including various completeness theorems. The book "Automation of Reasoning, Volumes 1 and 2, Classical Papers on Computational Logic", edited by Siekmann and Wrightson [**11**], contains the important research papers on the subject written before 1971. The editors have provided a valuable guide to this collection by designating which of the papers deserve special attention. The book "Automated Theorem Proving: After 25 Years", edited by Bledsoe and Loveland [**1**], contains a number of papers presented on the subject at the winter AMS meeting in 1983. It includes Loveland's paper [**6**], which reviews the previous 25 years of automated theorem proving—the field from which automated reasoning evolved—and the two papers given by the respective winners of the prize for milestones and the prize for current achievement in automated theorem proving.

With the intention of reaching a much wider audience than that aimed at by each of the preceding books, and to provide material not available elsewhere, a fifth book has just been completed. The book, "Automated Reasoning: Introduction and Applications" by Wos, Overbeek, Lusk, and Boyle [**19**], assumes no background. It covers various applications—mathematical, logical, and others—of the type discussed in this paper and, moreover, provides some guidance for using an automated reasoning program.

Very powerful and effective portable reasoning programs are now available [**7**], [**8**]. These programs, written in Pascal, can be run on relatively inexpensive computers, provided an appropriate Pascal compiler is accessible. They offer the researcher an arsenal of inference rules, strategies, and related processes from which to choose. Those who wish to experiment with and conduct research with an automated reasoning program can now do so.

**5. Amenable Questions.** An open question is amenable to attack with the assistance of an automated reasoning program if it is representable in the language of clauses. Being representable in first-order predicate calculus is sufficient, for an algorithm exists for mapping from that language to the language of clauses. Even better, an open question is amenable to such an attack if it can be represented by a finite set of equations. Specifically, if the structures under study are finitely based varieties—structures such as semigroups, groups, and rings—an automated reasoning program might well provide valuable assistance. Mappings of various kinds are perfectly acceptable. So also are finite sets of generators and relations. On the other hand, the statement "the structure is finite" cannot be represented in the language of clauses, and the fact that it is infinite can be represented but not in a useful fashion. To talk about "all mappings from one structure to another" is not possible, and to talk about "there exists a mapping" requires a trick. The trick is similar to one used in mathematics; we simply name the mapping in question. Problems in differential equations, for example, are not well suited to consideration. In short, problems similar to the ones discussed earlier are the best bet.

**6. Summary.** You are not required to be an expert in automated reasoning to use such a program. Its use is difficult, especially from the linguistic viewpoint. In fact, "the art of automated reasoning" captures the flavor of the present state of the discipline. Nevertheless, a number of activities are now benefiting from reliance on a reasoning program. They include research in mathematics and logic, the design and validation of logic circuits, and proving that computer programs fulfill certain claims made for them.

What is needed is additional open questions to attack with the assistance of a reasoning program. Such questions will lead to enhancements of existing techniques and the development of totally new techniques for problem solving. What is also needed is the use by mathematicians and logicians of automated reasoning programs. If open questions can be solved by researchers who know so little about the respective fields from which the problems were taken, how much more might be accomplished by experts in various fields with the assistance of such a program!

A reasoning program provides an explicit treatment of each proof and each model or counterexample it obtains. It offers a wide variety of inference rules, strategies, and related processes from which to choose when attacking a given problem. In view of the successes and properties cited here, perhaps mathematicians and logicians might strongly consider enlisting the aid of a different type of colleague—a computer program that functions as an automated reasoning assistant.

## References

**1.** W. W. Bledsoe and D. Loveland, eds., Automated Theorem Proving: After 25 Years, Contemporary Math., vol. 29, Providence, RI, 1984.

**2.** C. Chang and R. Lee, Symbolic Logic and Mechanical Theorem Proving, Academic Press, New York, 1973.

**3.** J. Guard, F. Oglesby, J. Bennett, and L. Settle, Semi-automated mathematics, J. Assoc. Comput. Mach., 16 (1969) 49–62.

**4.** R. Kowalski, Logic for Problem Solving, Elsevier, North-Holland, New York, 1979.

**5.** D. Loveland, Automated Theorem Proving: A Logical Basis, North-Holland, New York, 1978.

**6.** ______, Automated Theorem Proving: A Quarter Century Review, in Automated Theorem Proving: After 25 Years, ed. by W. W. Bledsoe and D. Loveland, Contemporary Math., vol. 29, pp. 1–45, Providence, RI, 1984.

**7.** E. Lusk, W. McCune, and R. Overbeek, Logic machine architecture: kernel functions, in 6th Conference on Automated Deduction, ed. by D. W. Loveland, Springer-Verlag, Lecture Notes in Computer Science, vol. 138, pp. 70–84, New York, 1982.

**8.** ______, Logic machine architecture: inference mechanism, in 6th Conference on Automated Deduction, ed. by D. W. Loveland, Springer-Verlag, Lecture Notes in Computer Science, vol. 138, pp. 85–108, New York, 1982.

**9.** J. McCharen, R. Overbeek, and L. Wos, Complexity and related enhancements for automated theorem-proving programs, Comput. Math. Appl., 2 (1976) 1–16.

**10.** J. Robinson, A machine-oriented logic based on the resolution principle, J. Assoc. Comput. Mach., 12 (1965) 23–41.

**11.** J. H. Siekmann and G. Wrightson, eds., The Automation of Reasoning, Vols. I and II, Springer-Verlag, Classical Papers on Computational Logic, New York, 1983.

**12.** B. Smith, Reference manual for the environmental theorem prover, an incarnation of AURA, Argonne National Laboratory Technical Report (to appear).

**13.** S. Winker, Generation and verification of finite models and counterexamples using an automated theorem prover answering two open questions, J. Assoc. Comput. Mach., 29 (1982) 273–284.

**14.** S. Winker and L. Wos, Automated generation of models and counterexamples and its application to open questions in ternary Boolean algebra, Proc. of the Eighth International Symposium on Multiple-valued Logic, IEEE and ACM, Rosemont, IL, 1978, pp. 251–256.

**15.** ______, New shortest single axioms for the equivalential calculus (in preparation).

**16.** S. Winker, L. Wos, and E. Lusk, Semigroups, antiautomorphisms, and involutions: A computer solution to an open problem, I, Math. Comp., 37 (1981) 533–545.

**17.** ______, Semigroups, antiautomorphisms, and involutions: A computer solution to an open problem, II (in preparation).

**18.** L. Wos, D. Carson, and G. Robinson, Efficiency and completeness of the set of support strategy in theorem proving, J. Assoc. Comput. Mach., 12, 4 (1965) 536–541.

**19.** L. Wos, R. Overbeek, E. Lusk, and J. Boyle, Automated Reasoning: Introduction and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1984.

**20.** L. Wos, G. Robinson, D. Carson and L. Shalla, The concept of demodulation in theorem proving, J. Assoc. Comput. Mach., 14 (1967) 698–709.

**21.** L. Wos, S. Winker, R. Veroff, B. Smith and L. Henschen, Questions concerning possible shortest single axioms for the equivalential calculus: An application of automated theorem proving to infinite domains, Notre Dame J. Formal Logic, 24, 2 (1983) 205–223.

Who is I. M. Milin? (See p. 130.)

**9.** M. Lerch, Dalši studie v oboru Malmsténovských řad, Rozpravy České Akad. 3, no. 28, 1894, 63 pp.

**10.** C. J. Moreno, The Chowla-Selberg formula, J. Number Theory, 17 (1983) 226–245.

**11.** T. Shintani, On special values of zeta-functions of totally real quadratic fields, Proc. Inter. Congr. Math., Helsinki, 1978, vol. 2, Acad. Sci. Fennica, Helsinki, 1980, pp. 591–597.

**12.** E. C. Titchmarsh, The Theory of the Riemann Zeta-Function, Clarendon Press, Oxford, 1951.

**13.** A. Weil, Elliptic Functions According to Eisenstein and Kronecker, Springer-Verlag, Berlin, 1976.

**14.** E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., University Press, Cambridge, 1962.


---

## 138. MISCELLANEA

Just as surely as our understanding of Nature is really valid only to the extent that it is mathematical, so also our understanding of higher domains must be based on mathematical models.

—R. Steiner (1861–1925), quoted in *Mathematical Reviews* 82k : 51001.

Each natural science contains genuine Science only to the extent that it contains Mathematics.... It may be that a formal philosophy of Nature in general, that is, a philosophy that deals only with general concepts, is possible without Mathematics, but a formal natural science dealing with definite objects (whether Physics or Psychology) is possible only with the use of Mathematics, and since each natural science contains only as much genuine Science as it contains *a priori* knowledge, it follows that a natural science is a genuine Science only to the extent that Mathematics can be applied to it.

—Immanuel Kant, Metaphysische Anfangsgründe der Naturwissenschaft, 1876.

*Perhaps we should sometimes listen to the other side. Goethe, for example, did not like Mathematics.*

Mathematicians are amazing people. In virtue of their accomplishments, they have set themselves up as a universal guild and will acknowledge only what suits their circle, what their method of organization can produce. A prominent mathematician once said, when someone strongly recommended a topic in Physics, "But can't it be reduced to calculation?"

—Goethe, Maximen und Reflexionen, no. 1277

It surely does not follow that the hunter who kills the game must also cook it. A cook might go hunting and shoot well; but he would be badly mistaken if he claimed that only a cook can be a good shot. It seems to me that this is the situation of mathematicians who claim that nobody can understand or discover physical phenomena without being a mathematician, since they should be pleased enough if the meat is brought to their kitchen for them to lard it with formulas and dress it as they like.

—Goethe, Maximen und Reflexionen, no. 1280

---

### ANSWER TO PHOTO ON PAGE 93

The photo is of Louis de Branges, who proved in 1984 the conjecture Bieberbach made in 1916, by establishing a strong inequality proposed by the Russian function-theorist I. M. Milin.

# SOLVABILITY OF INFINITE SYSTEMS OF POLYNOMIAL EQUATIONS OVER THE FIELD OF COMPLEX NUMBERS

ALEXANDER ABIAN
*Department of Mathematics, Iowa State University, Ames, IA 50011*

The purpose of this paper is to show that if every finite subsystem of a system of less than continuum many polynomial equations over the field of complex numbers has a solution, then the entire system has a solution. This is not the case over the field of real numbers even if the system is only countable, and it is not the case over the field of complex numbers if the system has continuum many equations.

It is known [**1**] that if every finite subsystem of an infinite (countable or uncountable) system $L$ of linear equations over a field $F$ has a solution in $F$, then $L$ has a solution in $F$. For systems of polynomial equations of higher degree the facts are different. For instance, over the field $R$ of real numbers, let us consider the countably infinite system of polynomial equations in the unknowns $x, y_1, y_2, y_3, \ldots$:

$$(1)\qquad \begin{aligned} (x-1) - y_1^2 &= 0 \\ (x-2) - y_2^2 &= 0 \\ (x-3) - y_3^2 &= 0 \\ \cdots &= \cdots \\ (x-m) - y_m^2 &= 0 \\ \cdots &= \cdots \end{aligned}$$

Every finite subsystem of (1) has a solution in $R$. To prove this, it is enough to show that, for every positive integer $m$, the finite subsystem of (1) consisting of the first $m$ equations of (1) has a solution, and that is so because

$$x = m, \quad y_1 = (m-1)^{1/2}, \quad y_2 = (m-2)^{1/2}, \quad y_3 = (m-3)^{1/2}, \ldots, \quad y_m = 0$$

gives a solution of the first $m$ equations of (1). The entire system (1), however, has no solution in $R$. Reason: if (1) had a solution in $R$ yielding, say, $x = r$, then for a positive integer $n > r$ the $n$th equation $(x-n) - y_n^2 = 0$ could not have a solution in $R$ since the square of a real number cannot be negative. This situation could not have happened over the field $C$ of complex numbers. Indeed,

$$x = 0, \quad y_1 = i, \quad y_2 = 2^{1/2}i, \ldots, \quad y_n = n^{1/2}i$$

is a solution in $C$ of the entire system (1).

On the other hand, over the field $C$ of complex numbers let us consider the system of continuum many polynomial equations given by:

$$(2)\qquad (z-c)y_c - 1 = 0 \qquad \text{with } c \in C.$$

Not only every finite subsystem of (2) but every proper subsystem of (2) has a solution in $C$. Indeed, if $P$ is a proper subsystem of (2) such that, say, for some $a$ in $C$ the equation

*Alexander Abian*: I received my M.S. from the University of Chicago in 1954 and my Ph.D. from the University of Cincinnati in 1956 under I. A. Barnett. I taught briefly at the University of Tennessee and at Queens College, NY, for three years at the University of Pennsylvania, then for five years at Ohio State University. Since 1967, I have been at Iowa State University. Since any nicely reasoned line (mathematical or otherwise) moves me deeply, I am almost equally attracted to the foundations of mathematics, algebra, topology, real, complex, functional and numerical analyses both in their pure and applied settings. This, I believe, is reflected in my various publications.

$(z-a)y_a - 1 = 0$ does not appear in $P$, then a solution of $P$ is given by $z = a$ and $y_c = (a - c)^{-1}$, which exists since $a \neq c$. Nevertheless, the entire system (2) has no solution in $C$. Reason: if (2) had a solution in $C$ yielding, say, $z = u$, then the equation $(z-u)y_u - 1 = 0$ would have no solution in $C$.

In contrast to the case of a countably infinite system of polynomial equations over the field of real numbers as well as a system of continuum many polynomial equations over the field of complex numbers, this paper shows that if a system of less than continuum many polynomial equations (each with finitely many unknowns, of course) such as for instance

$$\begin{aligned} a_1x^2 + a_2y^2z + a_3x^3z^5 + a_4y^2z^5 + \cdots + a_kx^5y^2z^3 + c_1 &= 0 \\ b_1z + b_2y^4z + b_3z^2v^5 + \cdots + b_myz^6v^2 + c_2 &= 0 \\ e_1xv^3 + e_2x^2uv^2 + e_3yz^3v^2 + e_4x^3v^4 + \cdots + e_nu^3v^4 + c_3 &= 0 \\ \cdots \qquad\qquad &= \cdots \end{aligned} \tag{3}$$

over the field $C$ of complex numbers is such that every finite subsystem of (3) has a solution in $C$, then the entire system (3) has a solution in $C$.

As expected and as (3) shows, by a *system of polynomial equations* we mean a system of equations where each equation is a polynomial in *finitely many* unknowns. The system may involve altogether an infinite number of unknowns.

In what follows $C$ stands for the field of complex numbers, all polynomials are over $C$, and all solutions are meant to be in $C$.

We need the following definitions:

A system $S$ of (polynomial) equations is called *consistent* if and only if every finite subsystem of $S$ has a solution.

Let $S$ be a system of (polynomial) equations and $z$ an unknown appearing in $S$, and let $c \in C$. The ordered pair $(z, c)$ is called an *assignment* of $z$ for the system $S$ if and only if the system obtained from $S$ by substituting $z = c$ consists of equations which are satisfied or which form a consistent system. Following tradition, instead of $(z, c)$ we shall refer to $z = c$ as an assignment of $z$.

Similarly, let $x, y, z, \ldots$ be unknowns which appear in a system $S$ of (polynomial) equations and let $p, q, c, \ldots$ be complex numbers. The set $\{(x,p),(y,q),(z,c),\ldots\}$ is called a *system of assignments* for $S$ if and only if the system obtained from $S$ by the substitutions $x = p$, $y = q$, $z = c, \ldots$ consists of equations which are satisfied or which form a consistent system.

Below we prove a nonobvious lemma of crucial importance; it is the key to what follows. It is not valid over the field of real numbers.

In what follows when we say that "an unknown *appears* in a polynomial equation $P(\ldots) = 0$" we mean that the coefficient of at least one term in which a nonzero power of that unknown occurs is not identically zero in $P(\ldots)$.

Also, in what follows by "*almost every*" we mean "*all but a finite number.*"

LEMMA 1. *If $P(x, \ldots, z) = 0$ is a polynomial equation in which the unknown $z$ appears, then either $z$ has only a finite number of assignments or else $z = c$ is an assignment of $z$ for almost every complex number $c$.*

*Proof*. If $z$ is the only unknown which appears in $P(x, \ldots, z) = 0$, then $P(x, \ldots, z)$ reduces to a nonconstant polynomial $P(z)$ in $z$ which (by the Fundamental Theorem of Algebra) has a finite number of zeros $c_1, \ldots, c_n$. Obviously, the finitely many assignments $z = c_1, \ldots, z = c_n$ are the only assignments of $z$. Hence in this case $z$ has only a finite number of assignments.

If besides $z$, finitely many other unknowns $x, y, \ldots, u$ also appear in $P(x, \ldots, z)$, then we rewrite $P(x, y, \ldots, u, z)$ as a polynomial in the descending powers of $x$, say, as:

$$x^kP_0(\ldots) + x^{k-1}P_1(\ldots) + \cdots + P_k(\ldots) = 0, \tag{4}$$

where the $P_i(\dots)$'s are polynomials in which some of the unknowns, $y, \dots, u, z$ may appear, and where $P_0(\dots)$ is *not identically zero*.

In connection with (4) one and only one of the following two cases must occur:

$$\text{in } P_0(\dots) \text{ the unknown } z \text{ does not appear} \tag{5}$$

or

$$\text{in } P_0(\dots) \text{ the unknown } z \text{ appears.} \tag{6}$$

Let (5) occur. Then, without loss of generality, we may assume that $P_0(\dots)$ is a polynomial $P_0(y, \dots, u)$ in which the unknowns $y, \dots, u$ appear. Since $P_0(y, \dots, u)$ is not identically zero, there exist complex numbers $p, \dots, q$ such that

$$P_0(p, \dots, q) \neq 0. \tag{7}$$

Assertion: $z = c$ is an assignment of $z$ for every complex number $c$. Indeed, since $z$ does not appear in $P_0(y, \dots, u)$, we see that upon substituting $y = p, \dots, u = q$ and $z = c$ for every complex number $c$ in $P(x, \dots, z)$ as given by (4), we get an equation that has a solution. This follows readily from (7) and from the Fundamental Theorem of Algebra. Hence in this case $z = c$ is an assignment of $z$ for every complex number $c$.

Let (6) occur. Again, without loss of generality, we may assume that $P_0(\dots)$ is a polynomial $P_0(y, \dots, u, z)$ in which the unknowns $y, \dots, u, z$ appear. Again, since $P_0(y, \dots, u, z)$ is not identically zero, there exist complex numbers $s, \dots, t$ such that

$$P_0(s, \dots, t, z) = P_0(z) \not\equiv 0. \tag{8}$$

Now, as a nonconstant polynomial in $z$, by the Fundamental Theorem of Algebra, $P_0(z)$ as given by (8) has finitely many zeros $h_1, \dots, h_m$. Thus

$$P_0(s, \dots, t, c) \neq 0 \text{ for every } c \text{ except } c = h_1, \dots, h_m$$

and consequently

$$P_0(s, \dots, t, z) \neq 0 \textit{ after the substitution } z = c, \textit{ for almost every } c \textit{ in } C. \tag{9}$$

We claim that $z = c$ is an assignment of $z$ for almost every complex number $c$. Indeed, we see that upon substituting $y = s, \dots, u = t$ and $z = c$ for almost every complex number $c$ in $P(x, \dots, z)$ as given by (4), we get a system that has a solution. This follows readily from (9) and from the Fundamental Theorem of Algebra. Hence in this case $z = c$ is an assignment of $z$ for almost every complex number $c$.

This completes the proof of Lemma 1.

REMARK 1. We observe that Lemma 1 is not valid over the field $R$ of real numbers. For instance, for the polynomial $z - y^2 = 0$ over $R$ there are continuum many assignments $z = r$ with $r \geqslant 0$ of $z$, but there are also continuum many nonassignments $z = r$ with $r < 0$ of $z$.

Now, we generalize Lemma 1 to the case of a finite system $F$ of consistent polynomial equations (i.e., to the case where the finite system $F$ is solvable). Our proof is based on Lemma 1 and on the existence of a resultant system of a finite system of polynomial equations [**2**, **3**].

LEMMA 2. *Let $F$ be a consistent finite system of polynomial equations in which the unknown $z$ appears. Then, for $F$, either $z$ has only a finite number of assignments or else $z = c$ is an assignment of $z$ for almost every complex number $c$.*

*Proof*. To the finite system $F$ we apply, say, Bezout's process of elimination of the unknowns appearing in $F$. We continue the process until it yields a finite system $L$ of resultant polynomial equations $L_1 = 0, \dots, L_n = 0$. We recall [**2**], [**3**] that $F$ is consistent (i.e., has a solution) if and only if $L$ is consistent (i.e., has a solution). Moreover, every solution of $L$ can be extended to a solution of $F$. Furthermore, no two distinct $L_1 = 0, \dots, L_n = 0$ have any unknown in common.

One and only one of the following two cases must occur:

(10) in $L$ the unknown $z$ does not appear

or

(11) in $L$ the unknown $z$ appears.

Let (10) occur. Since by the hypothesis of the Lemma $F$ is consistent, so is $L$ and since $z$ does not appear in $L$, obviously, $z = c$ is an assignment of $z$ (for $L$ as well as for $F$) for every complex number $c$.

Let (11) occur. Then the unknown $z$ appears in one and only one of the resultant polynomial equations, say, $L_1(x, \dots, z) = 0$. But then, by Lemma 1, (for $L$ as well as for $F$) the unknown $z$ has a finite number of assignments or else $z = c$ is an assignment of $z$ for almost every complex number $c$.

We extend Lemma 1 to the case of a consistent system $D$ of less than continuum many polynomial equations. Thus, every finite subsystem of $D$ has a solution and

$$\overline{\overline{D}} < 2^{\aleph_0} = \overline{\overline{C}} \tag{12}$$

(where $\overline{\overline{D}}$ is the cardinal number of the set $D$).

LEMMA 3. *Let $D$, with $\overline{\overline{D}} < \overline{\overline{C}}$, be a consistent system of polynomial equations in which the unknown $z$ appears. Then for $D$, either $z$ has a finite number of assignments or else $z = c$ is an assignment of $z$ for every complex number $c$ except for less than continuum many $c$'s.*

*Proof.* By Lemma 2, one and only one of the following two cases must occur:

(13) $D$ has a finite subsystem $F$ for which $z$ has only a finite number of assignments

or

(14) for every finite subsystem of $D$ it is the case that $z = c$ is an assignment of $z$ for almost every complex number $c$ (depending on the finite subsystem).

Let (13) occur and let $z = c_1, \dots, z = c_n$ be the finitely many assignments of $z$ for $F$. Let us assume to the contrary that none of these assignments is an assignment of $z$ for $D$. Thus, upon substituting $z = c_i$ (with $i = 1, \dots, n$) in $D$ we get a finite subsystem $F_i$ of $D$ which has no solution with $z = c_i$. Hence, the finite subsystem $F_1 \cup \cdots \cup F_n \cup \{F\}$ of $D$ has no solution whatsoever, and this contradicts the consistency of $D$. Thus, some of the finitely many $z = c_1, \dots, z = c_n$ must be assignments of $z$ for $D$ and since these are the only assignments of $z$ for $F$, we see that in this case $z$ has only a finite number of assignments for $D$.

Let (14) occur. Then $\overline{\overline{D}} < \overline{\overline{C}}$ by (12). Thus, $D$ has less than continuum many finite subsystems. We recall that less than continuum many finite subsets of complex numbers contains less than continuum many complex numbers. However, since there are continuum many complex numbers, from (14) it follows that $z = c$ is an assignment of $z$ for $D$ for every complex number $c$ except for less than continuum many $c$'s.

This completes the proof of Lemma 3.

REMARK 2. We observe that "for every except for less than continuum many" which appears in Lemma 3 cannot be replaced by "for almost every" which appears in Lemmas 1 and 2. We note also that Lemma 3 ceases to be valid if in it $D$ has continuum many polynomial equations (cf. our example (2)).

COROLLARY 1. *If $z$ is an unknown appearing in a consistent system $D$ of less than continuum many polynomial equations, then $z$ has an assignment for $D$.*

We recall our earlier definition of a system of assignments for a system of (polynomial) equations. Accordingly, in Corollary 1, if $(z, c)$ is an assignment, then $\{(z, c)\}$ is a system of assignment for $D$. Thus, we may rephrase Corollary 1 as follows.

COROLLARY 2. *The set of all the systems of assignments of a consistent system of less than continuum many polynomial equations is nonempty.*

Based on Corollary 2, we prove our main result:

THEOREM. *Let $D$ be a system of less than continuum many polynomial equations over the complex numbers $C$ such that every finite subsystem of $D$ has a solution in $C$. Then $D$ has a solution in $C$ (and conversely).*

*Proof.* Let $x, y, u, \ldots, z, \ldots$ be the unknowns which appear in $D$. By Corollary 2, the set $A$ of all the systems of assignments for $D$ is nonempty. Let

$$A = \{\{(x,a)\}, \{(x,b),(y,c)\}, \{(x,b),(y,t),(u,p)\}, \{(x,b),(y,t),(v,q),\ldots\},\ldots\}.$$

We partially order $A$ by the set theoretical inclusion $\subseteq$. It can be readily verified that every simply ordered subset of $(A, \subseteq)$ has a supremum. Thus, by Zorn's lemma, $(A, \subseteq)$ has a maximal element $M = \{(x,b),(y,t),(u,p),(v,q),\ldots\}$. To prove the theorem, it is enough to show that every unknown which appears in $D$ also appears in $M$. Let us assume that $z$ does not appear in $M$. Then upon substituting the elements of $M$, i.e., $x = b, y = t, u = p, v = q, \ldots$ in $D$, we get a system $D'$, with $\overline{\overline{D'}} < \overline{\overline{C}}$, which is a consistent system of polynomial equations in which the unknown $z$ appears. Thus, by Lemma 3 or Corollary 2, there exists an assignment $z = c$ of $z$ for $D'$. Clearly, $M \cup \{(z,c)\}$ would contradict the maximality of $M$ and therefore every unknown appearing in $D$ also appears in $M$. Obviously, $M$ yields a solution of $D$. (The converse is trivial.)

REMARK 3. Throughout this paper the field of complex numbers $C$ can be replaced by any algebraically closed field $A$, in which case $\overline{\overline{C}}$ must be replaced by $\overline{\overline{A}}$.

This work is partly supported by Iowa State University SHRI.

## References

1. A. Abian, Solvability of infinite systems of linear equations, Arch. Math., 12 (1976) 43–44.
2. B. L. van der Waerden, Modern Algebra, vol. 2, Ungar, NY, 1950, pp. 1–9.
3. W. V. Hodge and D. Pedoe, Methods of Algebraic Geometry, vol. 1, Cambridge Univ. Press, England, 1968, pp. 159–173.

---

## 137. MISCELLANEA

Two kinds of symbol must surely be distinguished. The algebraical symbol comes naked into the world of mathematics and is clothed with value by its masters. A poetic symbol—like the Rose, for Love, in Guillaume de Lorris—comes trailing clouds of glory from the real world, clouds whose shape and colour largely determine and explain its poetic use. In an equation, $x$ and $y$ will do as well as $a$ and $b$; but the *Romance of the Rose* could not, without loss, be re-written as the *Romance of the Onion*, and if a man did not see why, we could only send him back to the real world to study roses, onions, and love, all of them still untouched by poetry, still raw.

—E. M. W. Tillyard and C. S. Lewis,
*The Personal Heresy. A Controversy*,
Oxford University Press, London,
1939. (The excerpt above appears
in a chapter by C. S. Lewis.)

# THE LOGARITHMIC MEAN IN $n$ VARIABLES

A. O. PITTENGER

*Department of Mathematics and Computer Science, University of Maryland Baltimore County, Catonsville, MD 21228*

**1. Introduction.** In some recent notes in this MONTHLY, several authors have discussed properties of the logarithmic mean and variations thereof ([**2**], [**8**], [**10**], [**15**]). This little known "average" has been used to define a mean temperature between two points at different temperatures [**11**] and, somewhat surprisingly, in an analysis of index numbers in economics [**1**]. Further applications, however, require a "natural" generalization to a logarithmic mean of $n$ variables, and the purpose of this note is to discuss such an extension and to derive its basic properties.

As it happens, a suitable definition and some basic results are available in earlier papers of Carlson ([**3**], [**4**]). In Section 2 we present this definition and mention some of its properties. In Section 3 we let Jensen's inequality run rampant and establish in particular that the logarithmic mean in $n$ variables is bounded above and below by the arithmetic and geometric means, respectively. In fact we define an interpolating family $L_r$ of logarithmic means such that $L_{-n}$ is the geometric mean of $x_1, \ldots, x_n$, $L_{-1}$ is the logarithmic mean, $L_1$ is the arithmetic mean and $L_r$ is increasing in $r$ for fixed $x_1, \ldots, x_n$.

Applications may well require an explicit formula for the logarithmic mean, and that form is established in Section 4 for distinct $x_1, \ldots, x_n$. Other applications and, indeed, one of the motivations for considering this problem involve the asymptotic behavior of the logarithmic mean. It can be shown that under mild restrictions on the $x$'s, convergence of the arithmetic mean to a limit value implies the convergence of $L_r$ to that same value for all fixed $r$. In the context of probability theory, this becomes a strong law for $L_r$. Proofs of these more technical results will appear elsewhere.

**2. Definitions.** For positive $x$ and $y$ the logarithmic mean in two variables is defined as

$$L(x,y) = \begin{cases} \left( \dfrac{\log(y) - \log(x)}{y - x} \right)^{-1}, & x \neq y, \\ x, & x = y, \end{cases} \tag{2.1}$$

and some of the interest in this function stems from its relationship to the arithmetic and geometric means:

$$(x \cdot y)^{1/2} \leqslant L(x,y) \leqslant A_1(x,y) \equiv (x+y)/2. \tag{2.2}$$

In fact, if we define the $p$th arithmetic mean as

$$A_p(x,y) = \begin{cases} (xy)^{1/2}, & p = 0, \\ \left( \dfrac{x^p + y^p}{2} \right)^{1/p}, & p \neq 0, \end{cases} \tag{2.3}$$

then it is possible to show that the inequality

$$A_p(x,y) \leqslant L(x,y) \leqslant A_q(x,y) \tag{2.4}$$

---

*A. O. Pittenger*: I received my Ph.D. from Stanford in 1967, under the direction of K. L. Chung, and then held visiting or faculty positions at Moscow State University, the Rockefeller University, and the University of Michigan before settling down in Maryland in 1972. My primary area of research is in probability theory, but every now and then I am lured into problems involving elementary inequalities, a fact I have heretofore concealed from my probabilistic colleagues. My second profession is being the only non-musician in our family of five and fulfilling our motto: they play, I pay.

holds for all positive $x$ and $y$ if and only if $p \leqslant 0$ and $1/3 \leqslant q$. (These results were first obtained by Lin [10]. See also [13].)

The definition given in (2.1) does not suggest a reasonable generalization, although some authors have sought to find one based on that form [5]. However, following Carlson, we can write $\log(y) - \log(x)$ as $\int_x^y dt/t$ and make a change of variable to obtain the following felicitous form:

$$L(x, y) = \left(\int_0^1 \frac{dv}{vy + (1-v)x}\right)^{-1} \tag{2.5}$$

Thus $L$ involves the reciprocal of points in the convex hull of $x$ and $y$.

The representation (2.5) appears in Carlson [2] and is a special case of his multivariable definition which we now give. Let $A_{n-1}$ represent the simplex

$$A_{n-1} = \left\{(v_1, \dots, v_{n-1}) : 0 \leqslant v_i \leqslant 1, \quad \sum_1^{n-1} v_i \leqslant 1\right\}, \tag{2.6}$$

and let its volume be denoted by $m_{n-1}$. An easy calculation gives $m_{n-1} = ((n-1)!)^{-1}$. (As a point of interest, the volume is a special case of Dirichlet's integral. The reader unacquainted with that integral will find a nice presentation in [16, p. 258].

Letting $dv$ denote the differential of volume in $A_{n-1}$, we define $d\rho_{n-1}$ as the probability measure $(n-1)!dv$ and

$$L(x) = L(x_1, \dots, x_n) \equiv \left(\int_{A_{n-1}} (x \cdot v)^{-1} d\rho_{n-1}\right)^{-1}, \tag{2.7}$$

where

$$x \cdot v = \sum_{i=1}^n x_i v_i, \tag{2.8}$$

and $v_n = 1 - v_1 - \cdots - v_{n-1}$. In calculations below we will move from $x$ to $(x_1, \dots, x_n)$ without comment and will delete subscripts when no confusion will result.

Let us note that $L$ as defined in (2.7) does have some of the basic properties of a mean. It is obvious that

$$\min(x_1, \dots, x_n) \leqslant L(x) \leqslant \max(x_1, \dots, x_n),$$

and thus that if $\bar{x} = L(x_1, \dots, x_n)$, then $\bar{x} = L(\bar{x}, \dots, \bar{x})$. (However, unlike the multivariable arithmetic means, it is not true that $L(x_1, \dots, x_n)$ remains unchanged if $x_1, \dots, x_k$ are replaced by $L(x_1, \dots, x_k)$, for $k < n$.) Furthermore, $L$ is increasing with increasing $x$. Indeed if $0 < x_i \leqslant y_i$, then

$$\begin{aligned} 0 \leqslant L(y) - L(x) &= L(x) \cdot L(y) \cdot \int_A \frac{(y-x) \cdot v}{(x \cdot v)(y \cdot v)} d\rho \\ &\leqslant \frac{\left(\max_i (y_i)\right)^2}{\left(\min_i (x_i)\right)^2} \cdot \max(y_i - x_i). \end{aligned} \tag{2.9}$$

**3. Inequalities.** In Section 4 we compute an explicit form of $L$ which generalizes (2.1), but it is the integral representation that is more useful in establishing the theoretical properties of $L$. In fact, if we write

$$\phi(t) = t^{-1}, \tag{3.1}$$

the definition can be restated as

$$L(x) = \phi^{-1}\left(\int \phi(x \cdot v)\, d\rho\right), \tag{3.2}$$

a special case of Carlson's function $M$, [**4**, p. 33]. The form of (3.2) suggests the theory of means as presented in Hardy, Littlewood, and Pólya [**7**, p. 75] and we use their techniques below.

Since $\phi$ is convex and $\rho$ is a probability, Jensen's inequality implies that

$$\int \phi(x \cdot v)\, d\rho \geqslant \phi\left(\int (x \cdot v)\, d\rho\right).$$

Because $\phi$ is decreasing

$$\phi^{-1}\left(\int \phi(x \cdot v)\, d\rho\right) \leqslant \int (x \cdot v)\, d\rho = \sum x_i c_i,$$

where $c_i = \int v_i\, d\rho$. By symmetry the $c_i$ must all equal $n^{-1}$, and we have thus established the inequality with the arithmetic mean:

$$L(x_1, \dots, x_n) \leqslant A_1(x_1, \dots, x_n) = \sum x_i/n. \tag{3.3}$$

To obtain the geometric mean $(\Pi x_i)^{1/n}$ as a lower bound, we expand the scope of the discussion. Recall that as $p$ increases from zero to one, the $p$-arithmetic mean $A_p$ increases from the geometric mean $A_0$ to the arithmetic mean $A_1$. Such an interpolating family $L_r$ is defined by Carlson, and a further study is given in Stolarsky [**14**] for the case of two positive variables. In fact the problem settled for $L$ in (2.4) can be posed for $L_r$: find sharp upper and lower $p$-arithmetic mean bounds for $L_r(x, y)$. (That problem was settled in [**12**] and independently in [**15**]. See [**8**] for an overview.)

The definition of $L_r$ for two variables is

$$L_r(x, y) = \phi_r^{-1}\left(\int_0^1 \phi_r(vx + (1 - v)y)\, dv\right), \tag{3.4}$$

where

$$\phi_r(t) = \begin{cases} t^r, & r \neq 0, \\ \log(t), & r = 0. \end{cases} \tag{3.5}$$

The generalization is again immediate:

$$L_r(x) = \phi_r^{-1}\left(\int \phi_r(x \cdot v)\, d\rho\right). \tag{3.6}$$

Thus $L_{-1}(x) = L(x)$ and $L_1(x)$ is the arithmetic mean of $x_1, \dots, x_n$. The geometric mean can also be expressed in this notation.

(3.7) LEMMA. $(\Pi x_i)^{1/n} = L_{-n}(x_1, \dots, x_n)$.

*Proof*. This follows by induction. For $n = 2$ it is easy to integrate (3.4) and obtain $(xy)^{1/2} = L_{-2}(x, y)$. For the induction step we compute the inner-most integral in

$$L_{-(n+1)}^{-(n+1)}(x_1, \dots, x_{n+1}) = \int_0^1 n!\,dv_1 \int_0^{1-v_1} dv_2 \cdots \int_0^{1-v} dv_n \left(\sum x_i v_i\right)^{-n-1}$$

where $v = v_1 + \cdots + v_{n-1}$. If $x_n \neq x_{n+1}$, that computation gives

$$\frac{1}{n(x_n - x_{n+1})}\left\{\frac{1}{\left(\sum_1^{n-1} v_i x_i + (1 - v)x_{n+1}\right)^n} - \frac{1}{\left(\sum_1^{n-1} v_i x_i + (1 - v)x_n\right)^n}\right\}.$$

(If all the $x_i$ are equal, there's nothing to prove.) Using that expression in the remaining

$(n-1)$-fold integral and the induction hypothesis, we have

$$L_{-(n+1)}^{-(n+1)}(x_1,\ldots,x_{n+1}) = (x_n - x_{n+1})^{-1}\left\{\frac{1}{\prod_1^{n-1} x_i}\left(x_{n+1}^{-1} - x_n^{-1}\right)\right\}$$

$$= \left(\prod_1^{n+1} x_i\right)^{-1},$$

completing the proof.

We now have all the preliminary results for the following theorem, which is a special case of results given in [**3**] and [**4**].

(3.8) THEOREM. *If $r < s$, $L_r(x) \leqslant L_s(x)$. As a special case,*

$$(\Pi x_i)^{1/n} \leqslant L(x) \leqslant A_1(x), \tag{3.9}$$

*and equality holds if and only if the $x_i$ have a common value.*

*Proof.* The assertion is essentially a consequence of a discussion in Hardy, Littlewood, and Pólya [**7**, p. 75]. Assume $r < s$ and define

$$\Phi(t) = \phi_s \circ \phi_r^{-1}(t).$$

Thus

$$\Phi(t) = \begin{cases} t^{s/r}, & r < s, \text{ both nonzero,} \\ e^{st}, & r = 0 < s, \\ r^{-1}\log(t), & r < 0 = s. \end{cases}$$

It is easy to check that $\Phi$ is concave if $r < s < 0$ and is convex in all other cases. Hence, Jensen's inequality gives for suitable positive $f$

$$\Phi\left(\int f\,d\rho\right) \geqslant \int \Phi(f)\,d\rho, \qquad r < s < 0,$$

and

$$\Phi\left(\int f\,d\rho\right) \leqslant \int \Phi(f)\,d\rho, \qquad \text{otherwise.}$$

Using $f(v) = \phi_r(x \cdot v)$ and the structure of $\Phi$, we obtain

$$\phi_s(L_r) \geqslant \int \phi_s(x \cdot v)\,d\rho, \qquad \text{if } r < s < 0,$$

and

$$\phi_s(L_r) \leqslant \int \phi_s(x \cdot v)\,d\rho, \qquad \text{otherwise.}$$

If $s < 0$, $\phi_s^{-1}$ is decreasing, so that application of $\phi_s^{-1}$ reverses the inequality. In all other cases, $\phi_s^{-1}$ is increasing. Thus $L_r \leqslant L_s$ in all cases, and the assertion about strict inequality is a consequence of Jensen's inequality.

**4. An explicit form for L.** To use the logarithmic mean in applications, it would be nice to have an explicit formula for $L(x)$ which does not involve integration. Moreover, since we have identified the geometric mean of $x_1,\ldots,x_n$ as $L_{-n}(x)$, it is natural to attempt to find an explicit representation for $L_{-r}$ with $r$ an integer. In this section we give that representation for distinct $x_i$.

By the continuity of $L(x)$—and, analogously, of $L_{-r}(x)$—it is possible to obtain explicit formulae when two or more of the $x_i$ coincide.

First note that for $x_1 \neq x_2$

$$L^{-1}(x_1, x_2) = \frac{\log(x_2)}{x_2 - x_1} + \frac{\log(x_1)}{x_1 - x_2}.$$

Next, for distinct values of $x_1, x_2, x_3$, easy calculations give

$$L^{-1}(x_1, x_2, x_3) = 2 \sum_{i=1}^{3} \frac{x_i \log(x_i)}{\prod_{j \neq i} (x_i - x_j)}$$

and

$$L_{-2}^{-2}(x_1, x_2, x_3) = -2 \sum_{i=1}^{3} \frac{\log(x_i)}{\prod_{j \neq i} (x_i - x_j)}.$$

It is these results which lead us to the general form.

(4.1) PROPOSITION. *Let $p(i, 1, n) = \prod_{j=1, \neq i}^{n} (x_i - x_j)$. Then for distinct positive $x_i$ and integral $r$, $0 < r < n$,*

$$L_{-r}^{-r}(x_1, \ldots, x_n) = (-1)^{r-1} r \cdot \binom{n-1}{r} \sum_{i=1}^{n} \frac{x_i^{n-r-1} \log(x_i)}{p(i, 1, n)}.$$

*In particular, when $r = 1$ we have*

$$L^{-1}(x_1, \ldots, x_n) = (n-1) \sum_{i=1}^{n} \frac{x_i^{n-2} \log(x_i)}{p(i, 1, n)}.$$

In the proof of (4.1) we shall have occasion to use $p(i, 1, n)$ for different variables. When that occurs, the letter of the variables will appear. As an example, we have the following lemma whose proof we postpone.

(4.2) LEMMA. *Let $a_1, \ldots, a_s$ be $s$ distinct, nonzero real numbers. For $0 \leqslant k \leqslant s - 2$,*

$$\sum_{i=1}^{s} \frac{a_i^k}{p(i, 1, s; a)} = 0.$$

*Proof of* (4.1). The proof involves two inductions—the first on $r$ with $n = r + 1$ and the second on $n \geqslant r + 1$ with $r$ fixed. Both inductions tend to be notationally involved, and we omit many of the details.

For the first induction observe that for $r = 1$ and $n = 2$ the formula given by (4.1) coincides with the definition of $L^{-1}(x_1, x_2)$. For the induction step the trick is to integrate the innermost integral of $L_{-r-1}$, and that leads to

$$L_{-r-1}^{-r-1}(x_1, \ldots, x_{r+2}) = \frac{r+1}{r(x_{r+2} - x_{r+1})} \{ L_{-r}^{-r}(x_1, \ldots, x_{r+1}) - L_{-r}^{-r}(x_1, \ldots, x_r, x_{r+2}) \}.$$

We now use the induction hypothesis to replace each of the terms inside the brackets by a suitable summation, obtaining for the right-hand side of the foregoing the expression

$$\frac{(-1)^{r-1}(r+1)}{x_{r+2} - x_{r+1}} \left\{ \sum_{i=1}^{r} \log(x_i) \left( \frac{1}{p(i, 1, r+1)} - \frac{x_i - x_{r+1}}{p(i, 1, r+2} \right) \right.$$
$$\left. + \frac{\log(x_{r+1})}{p(r+1, 1, r)} - \frac{\log(x_{r+2})}{p(r+2, 1, r)} \right\}.$$

Some easy algebra allows this last expression to be written in the desired form.

Now for the second induction. For fixed $r$ we have just established (4.1) for $n = r + 1$. The trick needed here is to make a change of variables in the inner integrals by setting $v_i = (1 - v_1)u_i$ and $y_i = v_1 x_1 + (1 - v_1)x_i$ for $2 \leqslant i \leqslant n + 1$. Converting to the new variables leads to

$$L_{-r}^{-r}(x_1, \ldots, x_{n+1}) = \int_0^1 n(1 - v_1)^{n-1} L_{-r}^{-r}(y_2, \ldots, y_{n+1})\, dv_1,$$

and use of the induction hypothesis allows the right-hand side to be written as

$$n(-1)^{r-1} \cdot r\binom{n-1}{r} \sum_{i=2}^{n+1} \frac{1}{p(i,2,n+1;x)} \int_0^1 (a_i v_1 + x_i)^{n-r-1} \log(a_i v_1 + x_i)\, dv_1,$$

where $a_i = x_1 - x_i$.

It is easy to evaluate the definite integrals above, but when one tries to write the results in the form of (4.1) it becomes necessary to show that

$$\sum_{i=2}^{n+1} \frac{1}{p(i,1,n+1)} = \frac{-1}{p(i,1,n+1)} \tag{4.3}$$

and

$$\sum_{i=2}^{n+1} \frac{x_1^{n-r} - x_i^{n-r}}{p(i,1,n+1)} = 0. \tag{4.4}$$

Here's where we need Lemma (4.2): with $s = n + 1$ and $k$ chosen first as 0 and then as $n - r$, we can verify both (4.3) and (4.4). That step completes the second induction and the proof of (4.1).

*Proof of* (4.2). We have isolated (4.2) not only because it's required in the preceding proof, but also because it's a nice example of the technique of partial fractions. Under the hypotheses, we can write

$$\frac{x^{k+1}}{\prod_{i=1}^{s} (x - a_i)} = \sum_{i=1}^{s} \frac{A_i}{(x - a_i)}.$$

It is then trivial to verify that $A_i = a_i^{k+1}/p(i,1,s;a)$ and that (4.2) results from setting $x$ equal to zero.

**Acknowledgements.** It is a pleasure to thank I. A. Ahmad for rekindling my interest in logarithmic means, and K. S. Banerjee for insisting on an explicit form for $L$.

## References

1. K. S. Banerjee, On the existence of infinitely many log-change index numbers associated with the CES preference ordering, Statist. Hefte, Int. Zeitschrift f. Theorie u. Praxis, to appear.
2. B. C. Carlson, The logarithmic mean, this MONTHLY, 79 (1972) 615–618.
3. ______, A hypergeometric mean value, Proc. Amer. Math. Soc., 16 (1965) 1759–766.
4. ______, Some inequalities for hypergeometric functions, Proc. Amer. Math. Soc., 17 (1966) 32–39.
5. E. L. Dodd, Some generalizations of the logarithmic mean and of similar means of two variates which become indeterminate when the two variates are equal, Ann. Math. Stat., 12 (1971) 422–428.
6. W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, Wiley, New York, 1971.
7. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge Univ. Press, Cambridge, 1959.
8. E. B. Leach and M. C. Sholander, Extended mean values, this MONTHLY, 85 (1978) 84–90.
9. ______, Extended mean values II, J. Math. Anal. Appl., 92 (1983) 207–223.
10. T. P. Lin, The power mean and the logarithmic mean, this MONTHLY, 81 (1974) 879–883.
11. W. H. McAdams, Heat Transmission, McGraw-Hill, New York, 1954.
12. A. O. Pittenger, Inequalities between arithmetic and logarithmic means, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz., #678–715 (1980) 15–18.
13. ______, The symmetric, logarithmic and power means, ibid., 19–23.
14. K. B. Stolarsky, Generalizations of the logarithmic mean, Math. Mag., 48 (1975) 87–92.
15. ______, The power and generalized logarithmic means, this MONTHLY, 87 (1980) 545–548.
16. E. T. Whittaker and G. N. Watson, Modern Analysis, 4th ed., Cambridge Univ. Press, Cambridge, 1946.

# SYSTEMS OF NUMERATION

AVIEZRI S. FRAENKEL[1]

*Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4*

**1. Introduction.** There are many ways of representing an integer uniquely! The best known method is the decimal system. Whereas the Maya Indians used base 20 (using the 10 fingers and 10 toes), some of the human race became recently more primitive using the binary system instead, being influenced by the computer race which, for electronic reasons, is zealously addicted to the binary system. It may be of interest to electronic computers to know that there are actually infinitely many binary systems!

Somewhat less known systems of numeration include mixed radix, factorial representation, and exotic systems based on recurrence relations, a special case of which is the Fibonacci system of numeration. So there *are* many ways of representing an integer uniquely; many ways, that is, in each of which an integer can be represented uniquely.

These and other systems of numeration normally hide in various unexpected places, where they are applied for varied purposes. Typically, when the need for a numeration system arises, it is defined and an ad hoc proof of its capability to represent integers uniquely is given. The purpose of this article is to unify these results and show how they can be derived simply and uniformly.

In Theorem 1 we present a very simple yet general system of numeration. It may be used to derive all the numeration systems we intend to present, but some repetitive argumentation is involved. We prefer instead to use Theorem 1 to derive a general numeration system based on recursively defined basis elements. This is done in Theorem 2, which sheds more light on the nature of numeration systems than Theorem 1. We then derive our numeration systems from Theorem 2—all but one. The exceptional system is based on a recurrence relation with a negative coefficient, whereas the recurrence relations of Theorem 2 contain only positive coefficients. The exceptional system is therefore derived directly from Theorem 1.

Theorems 1 and 2 are given in Section 2. The derivation of the numeration systems from Theorem 2 is carried out in Section 3 in a rather neater way than Theorem 1 would permit. The exceptional system is derived in the final Section 4. Applications and uses of the numeration systems are briefly indicated. These include the ranking of permutations, of permutations with repetitions and of *Cayley-permutations*; polyphase sorting and merging of large data files, irregularities of distribution of sequences, the Chinese Remainder Theorem, various games, and a class of binary search trees called *cedar* trees. New applications motivated by this work include coding theory and the compression of sparse binary strings.

It should be pointed out that not all the known numeration systems can be derived from Theorem 1. An example is the *combinatorial numeration system*

$$N = \binom{a_n}{n} + \binom{a_{n-1}}{n-1} + \cdots + \binom{a_2}{2} + \binom{a_1}{1} \qquad (0 \leqslant a_1 < a_2 < \cdots < a_n)$$

(see, e.g., Lehmer [**12**]). There is a way of generalizing Theorem 1 (from numbers to infinite sets) so as to include also the combinatorial representation, but we prefer at this stage to keep our results as simple as possible. (For a fast algorithm for computing the "digits" $a_i$, see [**7**, Sect. 2].)

To keep the discussion simple, we state and prove our results for nonnegative integers only. At the end it will become clear that they hold, in fact, for any integer.

**2. Two basic numeration systems.** Let $1 = u_0 < u_1 < u_2 < \cdots$ be a finite or infinite sequence of integers. Let $N$ be any nonnegative integer, and suppose that $u_n$ is the largest number in the sequence not exceeding $N$ (except that we let $n = 0$ if $N = 0$). Dividing $N$ by $u_n$ and iterating

[1]Permanent address: Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel 76100. Work supported, in part, by Canadian Research Grant NSERC 69-0695.

For a biographical sketch of the author, see this MONTHLY, 89 (1982) 353.

gives

$$\begin{aligned}
N &= d_n u_n + r_n, & 0 &\leqslant r_n < u_n \\
r_n &= d_{n-1} u_{n-1} + r_{n-1}, & 0 &\leqslant r_{n-1} < u_{n-1} \\
r_{n-1} &= d_{n-2} u_{n-2} + r_{n-2}, & 0 &\leqslant r_{n-2} < u_{n-2} \\
&\vdots & &\vdots \\
r_{i+1} &= d_i u_i + r_i, & 0 &\leqslant r_i < u_i \\
&\vdots & &\vdots \\
r_2 &= d_1 u_1 + r_1, & 0 &\leqslant r_1 < u_1 \\
r_1 &= d_0 u_0. & &
\end{aligned}$$

Collecting terms we get

$$N = d_n u_n + d_{n-1} u_{n-1} + \cdots + d_0 u_0 \qquad (d_i \geqslant 0,\ i \geqslant 0).$$

This is the representation of $N$ in the numeration system $S = \{u_0, u_1, u_2, \dots\}$. The process above shows that every nonnegative integer can be represented in the system $S$. Note that

$$r_{i+1} = d_i u_i + d_{i-1} u_{i-1} + \cdots + d_0 u_0 < u_{i+1} \qquad (i \geqslant 0).$$

Conversely, if $N = \sum_{i=0}^{n} d_i u_i$, where

(1) $$d_i u_i + d_{i-1} u_{i-1} + \cdots + d_0 u_0 < u_{i+1} \qquad (i \geqslant 0),$$

then $\sum_{i=0}^{n} d_i u_i$ is the unique representation of $N$ by $S$, as shown in the following theorem:

THEOREM 1. *Let $1 = u_0 < u_1 < u_2 < \cdots$ be any finite or infinite sequence of integers. Any nonnegative integer $N$ has precisely one representation in the system $S = \{u_0, u_1, u_2, \dots\}$ of the form $N = \sum_{i=0}^{n} d_i u_i$, where the $d_i$ are nonnegative integers satisfying* (1).

*Proof.* It remains only to establish uniqueness. Suppose that $N$ has two representations:

$$N = c_n u_n + \cdots + c_0 u_0 = d_n u_n + \cdots + d_0 u_0,$$

where the digits $c_j$ and $d_j$ are nonnegative integers and satisfy (1). Let $i$ be the largest integer such that $c_{i+1} \neq d_{i+1}$, say $c_{i+1} > d_{i+1}$. Then

$$\begin{aligned}
u_{i+1} &\leqslant (c_{i+1} - d_{i+1}) u_{i+1} = (d_i - c_i) u_i + \cdots + (d_0 - c_0) u_0 \\
&\leqslant d_i u_i + \cdots + d_0 u_0,
\end{aligned}$$

contradicting (1).

The proof is complete by noting that the iterated division algorithm above produces a representation satisfying (1) for *every* nonnegative integer $N$. ■

REMARKS. (i) The existence of the representation has been shown in Yaglom and Yaglom [**19**, Chap. 8], where it is also stated that every $N$ has only one representation satisfying (1).

(ii) Since every nonnegative integer $N$ has a representation satisfying (1), it follows that if $N$ has a representation violating (1), then it has two representations. In this sense (1) is a necessary and sufficient condition for uniqueness.

(iii) Note that (1) implies

(2) $$0 \leqslant d_i < \frac{u_{i+1}}{u_i} \qquad (i \geqslant 0).$$

Now sometimes (2) implies (1) and sometimes it does not. When the $u_i$ are defined recursively, the situation depends on the length of the recurrence relation! If the recurrence relation contains only one term ($u_n = b^{(n)} u_{n-1}$), then (2) does imply (1). Therefore for the more conventional

numeration systems such as decimal, binary, mixed radix, and factorial systems, (2) is a necessary and sufficient condition for uniqueness. But for systems in which the recurrence relation contains more than one term, (2) is only a necessary but not a sufficient condition. This will become clear from the deeper Theorem 2 below.

For $m \geqslant 1$, let $b_1 = b_1^{(n)}, b_2, \dots, b_m$ be integers satisfying

$$1 \leqslant b_m \leqslant \cdots \leqslant b_2 \leqslant b_1^{(n)}$$

for all $n \geqslant 1$. Note that $b_2, \dots, b_m$ are constants, but $b_1 = b_1^{(n)}$ may depend on $n$. Let $u_{-m+1}, u_{-m+2}, \dots, u_{-1}$ be fixed nonnegative integers, and let

$$u_0 = 1,\ u_n = b_1^{(n)} u_{n-1} + b_2 u_{n-2} + \cdots + b_m u_{n-m} \qquad (n \geqslant 1), \tag{3}$$

be an increasing sequence.

Suppose that $N = \sum_{j=0}^{n} d_j u_j$ is the unique representation satisfying (1) of an integer $N$ in the system $S = \{u_0, u_1, u_2, \dots\}$, where the $u_i$ are defined by (3). Then the digits $d_i$ have to satisfy (2) by Remark (iii) above. If $m = 1$, (2) implies $d_i < u_{i+1}/u_i = b_1^{(i+1)}$ $(i \geqslant 0)$. If $m > 1$, then

$$\begin{aligned} d_i &< \left( b_1^{(i+1)} u_i + b_2 u_{i-1} + \cdots + b_m u_{i-m+1} \right) / u_i \\ &\leqslant b_1^{(i+1)} + \left( b_1^{(i)} u_{i-1} + b_2 u_{i-2} + \cdots + b_m u_{i-m} \right) / u_i = b_1^{(i+1)} + 1 \qquad (i \geqslant 1). \end{aligned}$$

By (3) (with $n = 1$), $d_0 < u_1 = \sum_{i=1}^{m} b_i u_{1-i}$, where $b_1 = b_1^{(1)}$. Thus

$$\begin{cases} 0 \leqslant d_i \leqslant b_1^{(i+1)} \quad (i \geqslant 1), \qquad 0 \leqslant d_0 < b_1^{(1)} + \sum_{j=2}^{m} b_j u_{1-j} \quad (m > 1), \\ 0 \leqslant d_i < b_1^{(i+1)} \quad (i \geqslant 0) \qquad (m = 1), \end{cases} \tag{4}$$

is a *necessary* condition for uniqueness. It is not in general sufficient, however. Here is the full—if slightly strange-looking—story.

THEOREM 2. *Let $S = \{u_i\}$ be a sequence of the form* (3). *Any nonnegative integer N has precisely one representation in S of the form $N = \sum_{i=0}^{n} d_i u_i$ if the digits $d_i$ are nonnegative integers satisfying the following (two-fold) condition*:

(i) *Let $k \geqslant m - 1$. For any j satisfying $0 \leqslant j \leqslant m - 2$, if*

$$\left( d_k, d_{k-1}, \dots, d_{k-j+1} \right) = \left( b_1^{(k+1)}, b_2, \dots, b_j \right), \tag{5}$$

*then $d_{k-j} \leqslant b_{j+1}$; and if* (5) *holds with $j = m - 1$, then $d_{k-m+1} < b_m$.*

(ii) *Let $0 \leqslant k < m - 1$. If* (5) *holds for any j satisfying $0 \leqslant j \leqslant k - 1$, then $d_{k-j} \leqslant b_{j+1}$; and if* (5) *holds with $j = k$, then $d_0 < \sum_{i=k+1}^{m} b_i u_{k+1-i}$.*

TABLE 1

| 5 | 1 | N | 18 | 5 | 1 | N | 236 | 65 | 18 | 5 | 1 | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 1 | | 2 | 2 | 12 | | | 2 | 0 | 0 | 36 |
| | 2 | 2 | | 2 | 3 | 13 | | | 2 | 0 | 1 | 37 |
| | 3 | 3 | | 2 | 4 | 14 | | | | | | ⋮ |
| | 4 | 4 | | 3 | 0 | 15 | | | 3 | 1 | 4 | 63 |
| 1 | 0 | 5 | | 3 | 1 | 16 | | | 3 | 2 | 0 | 64 |
| 1 | 1 | 6 | | 3 | 2 | 17 | | 1 | 0 | 0 | 0 | 65 |
| 1 | 2 | 7 | 1 | 0 | 0 | 18 | | 1 | 0 | 0 | 1 | 66 |
| 1 | 3 | 8 | 1 | 0 | 1 | 19 | | | | | | ⋮ |
| 1 | 4 | 9 | | | | ⋮ | | 3 | 2 | 0 | 4 | 235 |
| 2 | 0 | 10 | 1 | 3 | 1 | 34 | 1 | 0 | 0 | 0 | 0 | 236 |
| 2 | 1 | 11 | 1 | 3 | 2 | 35 | 1 | 0 | 0 | 0 | 1 | 237 |

EXAMPLE. Let us examine the numeration system defined by $m = 3$, $u_{-2} = 0$, $u_{-1} = u_0 = 1$, $u_n = 3u_{n-1} + 2u_{n-2} + u_{n-3}$ $(n \geqslant 1)$. Thus $b_1 = 3$, $b_2 = 2$, $b_3 = 1$, and $S =$

$\{1,5,18,65,236,\dots\}$. By (4), the digit bounds are $0 \leqslant d_i \leqslant 3$ $(i \geqslant 1)$ and $0 \leqslant d_0 \leqslant 4$. Table 1 exhibits the representation by $S$ of a few integers. We point out, in particular, that by subcondition (i) the representation of 65 is 1000 rather than 321 as might be expected at first, and the representation of 236 is 10000 rather than 3210. By subcondition (ii) the representation of 18 is 100 rather than 33, and the representation of 36 is 200 rather than 133.

Note that subconditions (i) and (ii) are both concerned with blocks of consecutive digits. They differ only in the location of these blocks: in (i) the right-hand digit $d_{k-m+1}$ of a block of maximal size $m$ is in some position $k-m+1 \geqslant 0$. Such a block may appear anywhere within the digit sequence. If it is "shifted right" until its rightmost digit goes beyond position 0, we get a truncated block: the right-hand digit of the truncated block (of smaller maximal size $k+1<m$) occupies position 0. This is the situation considered in (ii).

Further note that (4) is implied by the condition of the theorem: the first inequality of the case $m>1$ of (4) is obtained by putting $j=0$ in the first part of (i); the second by putting $j=k=0$ in the second part of (ii). For $m=1$, (i) with $j=m-1$ gives $0 \leqslant d_k < b_1^{(k+1)}$ $(k \geqslant 0)$ and (ii) is empty. Also the first part of (i) is empty. So for $m=1$ the condition is equivalent to (4).

To motivate further the somewhat curious condition of Theorem 2, consider the violation of (i) in the special case when (5) holds for $j=m-1$ but $d_{k-m+1} \geqslant b_m$. Then the integer $M=\sum_{i=k-m+1}^{k} d_i u_i$ has two representations:

$$M = b_1^{(k+1)} u_k + \sum_{i=k-m+2}^{k-1} b_{k-i+1} u_i + d_{k-m+1} u_{k-m+1},$$

and $M = u_{k+1} + (d_{k-m+1} - b_m) u_{k-m+1}$. The latter, which violates (1), follows from the former representation using the recurrence (3). (After Lemma 1 below, it may be seen that also the other violations of the condition of Theorem 2 boil down to the case of a sequence of digits sufficiently large, so that a number has an *additional* representation, violating (1), obtained by employing the recurrence (3).)

We finally remark that Theorem 2 does not consider the most general case (for example, some negative coefficients could be permitted in the recurrence relation), but it suffices for deriving in a simple manner all but one of the numeration systems of interest to us.

*Proof of Theorem* 2. We first show that a representation satisfying the condition of the theorem always exists. Let $N$ be any nonnegative integer and let $N=\sum_{i=0}^{n} d_i u_i$ be its representation as produced by the iterated division algorithm preceding Theorem 1. Assume that (i) does not hold. Suppose first that there is some $j$ satisfying $1 \leqslant j \leqslant m-2$ for which (5) holds but $d_{k-j} > b_{j+1}$. Then

$$\begin{aligned}\sum_{i=0}^{k} d_i u_i \geqslant \sum_{i=k-j}^{k} d_i u_i &\geqslant b_1^{(k+1)} u_k + \sum_{i=2}^{j+1} b_i u_{k+1-i} + u_{k-j}\\ &= b_1^{(k+1)} u_k + \sum_{i=2}^{j+1} b_i u_{k+1-i} + b_1^{(k-j)} u_{k-j-1} + \sum_{i=2}^{m} b_i u_{k-j-i}\\ &\geqslant b_1^{(k+1)} u_k + \sum_{i=2}^{m} b_i u_{k+1-i} = u_{k+1},\end{aligned}$$

violating (1). Secondly suppose that (5) holds with $j=m-1$ but $d_{k-m+1} \geqslant b_m$. Then

$$\sum_{i=0}^{k} d_i u_i \geqslant \sum_{i=k+1-m}^{k} d_i u_i \geqslant b_1^{(k+1)} u_k + \sum_{i=2}^{m} b_i u_{k+1-i} = u_{k+1},$$

again violating (1). If we assume that (ii) does not hold, then the same arguments show again that (1) is violated.

It follows that the representation produced by the iterated division algorithm satisfies the

condition of Theorem 2—in addition to (1).

For demonstrating that there is only one representation satisfying the condition, we first prove an auxiliary result which sheds a light of its own on the nature of general numeration systems.

LEMMA 1. (i) *For any nonnegative integer* $t$, *write* $t+1 = qm + r$, $0 \leqslant r < m$. *Consider the following representation of a nonnegative integer* $f(t)$ *in the system* (3):

$$\begin{aligned} f(t) = {} & b_1^{(t+1)} u_t + b_2 u_{t-1} + \cdots + b_{m-1} u_{t+2-m} + (b_m - 1) u_{t+1-m} \\ & + b_1^{(t+1-m)} u_{t-m} + b_2 u_{t-1-m} + \cdots + b_{m-1} u_{t+2-2m} + (b_m - 1) u_{t+1-2m} \\ & + \cdots \\ & + b_1^{(t+1-(q-1)m)} u_{t-(q-1)m} + b_2 u_{t-1-(q-1)m} + \cdots + b_{m-1} u_{t+2-qm} + (b_m - 1) u_{t+1-qm} \\ & + b_1^{(t+1-qm)} u_{t-qm} + b_2 u_{t-1-qm} + \cdots + b_{r-1} u_{t+2-r-qm} + d_0 u_0, \end{aligned}$$

*where* $d_0 = b_m - 1$ (*if* $r = 0$); $d_0 = \sum_{j=0}^{m-r} b_{r+j} u_{-j} - 1$ (*if* $r > 0$). *Then* $f(t) = u_{t+1} - 1$.

(ii) *Let* $N_1 = \sum_{i=0}^{n} d_i u_i$ *and* $N_2 = \sum_{i=l+1}^{n} d_i u_i + \sum_{i=0}^{l} c_i u_i$, *where* $c_l < d_l$, $l$ *any integer in* $[0, n]$, *and the* $c_i, d_i$ *are arbitrary digits satisfying the condition of Theorem* 2. *Then* $N_2 < N_1$.

Part (i) of the lemma states that $f(t)$ is the analog of $99\ldots9$ ($t+1$ 9's) in decimal. Part (ii) states that if two representations differ in the $l$th position being identical to the left of it, then the representation with the larger digit in position $l$ corresponds to the larger number. This is of course trivial for, say, the decimal system. In the present more general case, however, this result is not a priori clear: Suppose for example that—for $m > 3$—three consecutive digits of some representation of some number satisfy $(d_k, d_{k-1}, d_{k-2}) = (b_1^{(k+1)}, b_2, b_3)$. By the condition of Theorem 2 these digits are maximal, in the sense that none of them can be replaced by a larger one. However, replacing, say, $d_{k-1} = b_2$ by $b_2 - 1$ enables replacing $d_{k-2} = b_3$ by $d_{k-2} = b_1^{(k-1)}$, which may be considerably larger than $b_3$. Part (ii) states that such a replacement yields a smaller, not a larger number as might be expected.

*Proof of Lemma* 1. It is straightforward to verify the following two facts:

(a) The representation of $f(t)$ as given in (i) satisfies the condition of Theorem 2.

(b) Increasing *any* digit of this representation while leaving the digits to its left unchanged results in a representation violating the condition of Theorem 2.

For the proof of (i), note that (3) implies directly

$$\begin{aligned} f(t) = {} & (u_{t+1} - u_{t+1-m}) + (u_{t+1-m} - u_{t+1-2m}) + \cdots \\ & + (u_{t+1-(q-1)m} - u_{t+1-qm}) + (u_{t+1-qm} - 1) = u_{t+1} - 1. \end{aligned}$$

We prove (ii) by induction on the "digital position" $l$. The result is clear for $l = 0$. Suppose the result holds for any two representations satisfying the condition of Theorem 2 whose leftmost *differing* digits are in some position $\leqslant l$. We have to show that, given any representation $N_1 = \sum_{i=0}^{n} d_i u_i$, we have $N_3 < N_1$, where $N_3 = \sum_{i=l+2}^{n} d_i u_i + \sum_{i=0}^{l+1} g_i u_i$, $g_{l+1} < d_{l+1}$, and the representations of $N_1$ and $N_3$ satisfy the condition of Theorem 2. Subtracting, we get

$$N_1 - N_3 = (d_{l+1} - g_{l+1}) u_{l+1} + \sum_{i=0}^{l} (d_i - g_i) u_i \geqslant u_{l+1} - \sum_{i=0}^{l} g_i u_i.$$

Let $g(l) = \sum_{i=0}^{l} g_i u_i$. If the representations of $f(l) = \sum_{i=0}^{l} f_i u_i$ as given in (i) of the lemma (with $t$ replaced by $l$) and $g(l)$ disagree, then the leftmost disagreement occurs in position $k$ for some $k \in [0, l]$, that is, $f_k \neq g_k$, but $f_i = g_i$ for $k < i \leqslant l$. Since the representation $g(l) = \sum_{i=0}^{l} g_i u_i$ satisfies the condition of Theorem 2, fact (b) implies $g_k < f_k$. By our induction hypothesis we thus have $g(l) \leqslant f(l)$. Since by (i) $f(l) = u_{l+1} - 1$, we thus have $N_1 - N_3 \geqslant 1$; that is $N_3 < N_1$. ■

We now resume the proof of Theorem 2. Let $N = \sum_{i=0}^{n} d_i u_i$ be a representation satisfying the condition of Theorem 2. If the representations of $f(n)$ and $N$ are different, then fact (b) implies

$d_k < f_k$ for some $k \in [0, n]$. By (ii) and (i) of Lemma 1, we thus have $N \leqslant f(n) < u_{n+1}$, so inequality (1) is satisfied. Hence the representation $\Sigma_{i=0}^{n} d_i u_i$ of $N$ is unique by Theorem 1. ■

We note that the proof above demonstrated the equivalence of (1) and the condition of Theorem 2 for the system $\{u_i\}$ of the form (3). In particular, if $N$ has a representation violating the condition, then $N$ has two representations, since it also has a representation satisfying the condition. Analogously to remark (ii) above we can thus say that the condition of Theorem 2 is a necessary and sufficient condition for uniqueness. In the applications below we will see the form this condition assumes for each case.

**3. A spectrum of numeration systems.** We shall now use Theorem 2 to derive several families of useful numeration systems. Existence of these families is evident from the iterated division algorithm just preceding Theorem 1. It will therefore suffice to demonstrate the digit bounds required for uniqueness. Recall that $u_0 = 1$ for all systems.

*Polynomial systems.* Let $b > 1$ be a fixed integer and let $u_n = b^n$, that is, $u_{n+1} = bu_n$ $(n \geqslant 0)$. Let $N$ be any nonnegative integer. Since the recurrence for the $u_i$ has length 1, Theorem 2 implies that the representation $N = \Sigma_{i=0}^{n} d_i b^i$ is unique if and only if $0 \leqslant d_i < b$ $(i \geqslant 0)$. This gives the most commonly used numeration systems, such as the decimal $(b = 10)$ and the binary $(b = 2)$ system.

*Mixed radix.* Let $1 = a_0, a_1, a_2, \ldots$ be any sequence of integers with $a_i > 1$ $(i \geqslant 1)$, and let $u_n = a_0 a_1 \ldots a_n$, that is, $u_{n+1} = a_{n+1} u_n$ $(n \geqslant 0)$. By the argument above, the representation $N = \Sigma_{i=0}^{n} d_i a_0 \ldots a_i$ is unique if and only if $0 \leqslant d_i < a_{i+1}$ $(i \geqslant 0)$. The mixed radix representation has been used for a constructive proof of the generalized Chinese Remainder Theorem (see [**3**] and Knuth [**10**, Sect. 4.3.2]), and, in conjunction with other numeration systems, for ranking permutations with repetitions and *Cayley-permutations* [**15**]. The latter method has been applied for compressing and partitioning large dictionaries in order to enable the storage of their "information bearing" parts in high-speed memory [**7**].

*Factorial representation.* This is the special case of the mixed radix representation where $a_n = n + 1$, leading to $u_n = (n+1)!$ $(n \geqslant 0)$. Thus the representation $N = \Sigma_{i=0}^{n} d_i (i+1)!$ is unique if and only if $d_i \leqslant i + 1$ $(i \geqslant 0)$. The factorial representation has been used for ranking permutations; see Lehmer [**12**] and Even [**2**, Chap. 1].

*Reflected factorial representation.* To represent a nonnegative integer $N$, select $h$ with $h! > N$, and let $u_n = h!/(h-n)!$, that is, $u_{n+1} = u_n(h-n)$ $(n \geqslant 0)$. Since again the recurrence has length 1 only, the representation $N = \Sigma_{i=0}^{h-2} d_i h!/(h-i)!$ is unique if and only if $0 \leqslant d_i < h - i$ $(0 \leqslant i \leqslant h-2)$. The reflected factorial representation has also been used for ranking permutations [**2**, Chap. 1].

Up to this point all systems used only a one-term recurrence relation for the $u_i$ (the case $m = 1$ in Theorem 2). This produced the better known numeration systems. The more exotic systems are obtained for $m > 1$. In these cases requirement (4) does not suffice to insure uniqueness, and the condition of Theorem 2 is needed to guarantee it. We start with an example illustrating the case $m = 2$.

*Continued fraction representation.* Let $\alpha$ be an irrational number satisfying $1 < \alpha < 2$. Then $\alpha$ has a unique *simple continued fraction* expansion of the form

$$\alpha = 1 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}} = [1, a_1, a_2, a_3, \ldots],$$

where the $a_i$ are positive integers. Its *convergents* $p_n/q_n = [1, a_1, a_2, \ldots, a_n]$ satisfy the recursion

$$p_{-1} = 1,\ p_0 = 1,\ p_n = a_n p_{n-1} + p_{n-2} \qquad (n \geqslant 1),$$
$$q_{-1} = 0,\ q_0 = 1,\ q_n = a_n q_{n-1} + q_{n-2} \qquad (n \geqslant 1).$$

See, e.g., Hardy and Wright [**8**, Chap. 10], Olds [**16**], or Perron [**17**].

THEOREM 3. *Every nonnegative integer has precisely one representation of the form*

$$N = \sum_{i=0}^{k} s_i p_i,\ 0 \leqslant s_i \leqslant a_{i+1};\ s_{i+1} = a_{i+2} \Rightarrow s_i = 0 \qquad (i \geqslant 0), \tag{6}$$

*and also precisely one representation of the form*

$$N = \sum_{i=0}^{l} t_i q_i,\ 0 \leqslant t_0 < a_1,\ 0 \leqslant t_i \leqslant a_{i+1};\ t_i = a_{i+1} \Rightarrow t_{i-1} = 0 \qquad (i \geqslant 1). \tag{7}$$

*Proof.* The requirements (4) imply the bounds on the digits $s_i$ and $t_i$; and the condition of Theorem 2 implies that if one of these digits attains its maximal value, then its right-hand-side neighbor must vanish. ∎

We remark again that if any of the digit bounds in (6) or (7) is violated, then the condition of Theorem 2 is violated. Since $N$ has also a representation satisfying the condition, $N$ has two representation in this case. Thus (6) and (7) are necessary and sufficient for uniqueness in the two numeration systems.

If $a_i = 1$ $(i \geqslant 1)$, then $\alpha = [\dot{1}] = (1 + \sqrt{5})/2$ is the golden ratio (where $\dot{x}$ denotes the infinite concatenation of $x$ with itself). In this case the system (6) becomes the *Fibonacci numeration system* which is a binary system (digits 0 and 1 only), with the proviso that two adjacent 1's never occur. See Zeckendorf [**20**]. This system lies behind the Fibonacci search (see Knuth [**11**, Sect. 6.2.1]); it has also been used by Whinihan [**18**], [**9**, Sect. 1.2.8] for giving the winning strategy of a game on a pile of tokens.

Numeration systems of the form (6) and (7) can also be defined for rational $\alpha$. An interesting relationship exists between $n$ expressed in the system (7) and $\lfloor n\alpha \rfloor$ and $\lfloor n\beta \rfloor$ in the system (6), where $a^{-1} + \beta^{-1} = 1$ [**6**]. This relationship is particularly interesting for the special case $\alpha = [1, \dot{a}]$, where $a$ is any positive integer. It can be utilized for giving a winning strategy for generalized Wythoff games both in normal play [**4**] and in misère play [**5**]. The class of *cedar trees* consolidates the winning strategies of these games. The case $a = 1$ gives a strategy for the classical Wythoff game [**19**], which is a two-player game played with two piles of tokens. Each player may remove any number of tokens from a single pile or the same number of tokens from both piles. In normal play, the player first unable to move is the loser, his opponent the winner. This outcome convention is reversed for misère play.

An interesting class of numeration systems containing those considered in Theorem 3 is induced by expanding $\alpha$ in a not necessarily simple continued fraction:

$$\alpha = 1 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cfrac{b_4}{\ddots}}}}.$$

We shall not explore this family here any further, however.

We consider next an example of an arbitrary length recurrence relation.

*Higher order Fibonacci systems.* Fibonacci numbers of order $m$ $(m \geqslant 2)$ are defined by

$$u_{-m+1} = u_{-m+2} = \cdots = u_{-2} = 0,\ u_{-1} = u_0 = 1,$$
$$u_n = u_{n-1} + u_{n-2} + \cdots + u_{n-m} \qquad (n \geqslant 1).$$

(For simplicity we write $u_i$ instead of $u_i^{(m)}$.) This definition gives the ordinary Fibonacci numbers for $m = 2$.

It follows directly from Theorem 2 that any nonnegative integer $N$ has precisely one binary representation of the form

$$N = \sum_{i=0}^{n} d_i u_i \qquad (0 \leqslant d_i \leqslant 1, 0 \leqslant i \leqslant n),$$

such that there is no run of $m$ consecutive 1's. Because if there is no such run, then the condition of Theorem 2 is satisfied and the representation is therefore unique; if the representation of $N$ contains a run of $m$ consecutive 1's, then the condition of Theorem 2 is violated. By Theorem 2, $N$ has also a representation without a run of $m$ consecutive 1's, so $N$ has two representations in this case.

Since an $m$-order Fibonacci system exists for every $m \geqslant 2$, there *are* infinitely many binary systems as claimed at the beginning of the article. In a similar way we can in fact construct infinitely many $b$-ary systems for every integer $b \geqslant 2$.

Fibonacci numbers of order $m$ have been used by Knuth [**11**, Sect. 5.4.2] for polyphase merge of data runs stored on magnetic tape transports: we like to sort-merge data runs stored on $m + 1$ magnetic tape transports such that at each stage $m$ tapes merge into one tape and so that the tapes run continuously, to save time. This can be done if at each stage the number of runs on one of the tapes is an $m$th order Fibonacci number, and a simple function thereof on the other tapes. The same numeration system but with different values for $u_{-m+1}, u_{-m+2}, \ldots, u_{-1}$ has been applied by Lynch [**13**] to polyphase sorting.

**4. Another continued fraction system.** Let $\alpha = [1, a_1, a, a_3, a, \ldots]$, that is $a_{2n} = a \quad (n \geqslant 1)$, where $a$ is any positive integer. Further, let $u_n$ stand for either $p_n$ or $q_n$, the understanding being that in each formula involving $u_i$, either all $u_i$ stand for $p_i$ or all stand for $q_i$. We shall develop two numeration systems based on the numerators and denominators of the even convergents of $\alpha$. Let us start with two auxiliary results on the even convergents.

$$\text{Throughout we let } \varepsilon = \begin{cases} 0, & \text{if } u_i = q_i, \\ 1, & \text{if } u_i = p_i \end{cases} \quad (i \geqslant -1).$$

LEMMA 2. *If we define $u_{-2}$ to be $1 - \varepsilon a$ (and $u_0 = 1$), then the even convergents of $\alpha$ satisfy*

$$u_{2n} = (aa_{2n-1} + 2)u_{2n-2} - u_{2n-4} \qquad (n \geqslant 1).$$

*Proof.* For $n \geqslant 1$ we have,

$$\begin{aligned} u_{2n} &= au_{2n-1} + u_{2n-2} = a(a_{2n-1}u_{2n-2} + u_{2n-3}) + u_{2n-2} \\ &= aa_{2n-1}u_{2n-2} + (u_{2n-2} - u_{2n-4}) + u_{2n-2} \\ &= (aa_{2n-1} + 2)u_{2n-2} - u_{2n-4}. \blacksquare \end{aligned}$$

LEMMA 3. *Let $0 \leqslant k < l$. Then*

$$\begin{aligned} u_{2l+2} = {} & a(a_{2l+1}u_{2l} + a_{2l-1}u_{2l-2} + \cdots + a_{2k+3}u_{2k+2} + a_{2k+1}u_{2k}) \\ & + u_{2l} + u_{2k} - u_{2k-2}. \end{aligned}$$

*Proof.* By Lemma 2,

$$\begin{aligned} u_{2l+2} &= (aa_{2l+1} + 2)u_{2l} - u_{2l-2} \\ u_{2l} &= (aa_{2l-1} + 2)u_{2l-2} - u_{2l-4} \\ u_{2l-2} &= (aa_{2l-3} + 2)u_{2l-4} - u_{2l-6} \\ &\vdots \end{aligned}$$

$$u_{2k+6} = (aa_{2k+5} + 2)u_{2k+4} - u_{2k+2}$$
$$u_{2k+4} = (aa_{2k+3} + 2)u_{2k+2} - u_{2k}$$
$$u_{2k+2} = (aa_{2k+1} + 2)u_{2k} - u_{2k-2}.$$

Adding, we get the claimed result. ■

We are now ready to present the last family of numeration systems considered here.

THEOREM 4. *Every nonnegative integer has precisely one representation of the form* $N = \sum_{i=0}^{n} d_{2i}u_{2i}$, *where the digits* $d_{2i}$ *are nonnegative integers satisfying*

$$(8) \qquad 0 \leqslant d_{2i} \leqslant aa_{2i+1} + 1 \quad (i \geqslant 1), \qquad 0 \leqslant d_0 \leqslant a(a_1 + \varepsilon)$$

*and the additional condition: If for some* $0 \leqslant k < l \leqslant n$, $d_{2l}$ *and* $d_{2k}$ *attain their maximal values, then there exists j satisfying* $k < j < l$ *(so actually* $l - k \geqslant 2$*) such that* $d_{2j} < aa_{2j+1}$.

*Proof.* Let $n$ be any nonnegative integer. To show that $N$ has a representation of the required form, apply the iterated division algorithm to produce the representation $N = \sum_{i=0}^{n} d_i u_i$. This representation satisfies (1) and (2). By (2),

$$d_{2i} < \frac{u_{2i+2}}{u_{2i}} = \frac{(aa_{2i+1} + 2)u_{2i} - u_{2i-2}}{u_{2i}} = aa_{2i+1} + 2 - \frac{u_{2i-2}}{u_{2i}} \qquad (i \geqslant 0),$$

which implies the bounds (8). Suppose the additional condition does not hold. If $k > 0$, then

$$\sum_{i=0}^{l} d_{2i}u_{2i} \geqslant \sum_{i=k}^{l} d_{2i}u_{2i} \geqslant a(a_{2l+1}u_{2l} + \cdots + a_{2k+1}u_{2k})$$
$$+ u_{2l} + u_{2k} = u_{2l+2} + u_{2k-2} > u_{2l+2},$$

by Lemma 3. This violates (1). If $k = 0$, then

$$\sum_{i=0}^{l} d_{2i}u_{2i} \geqslant a(a_{2l+1}u_{2l} + \cdots + a_3u_2 + a_1u_0)$$
$$+ u_{2l} + \varepsilon a u_0 = u_{2l+2} + u_{-2} + (\varepsilon a - 1)u_0 = u_{2l+2},$$

again violating (1). It follows that any representation satisfying (1) also satisfies (8) and the additional condition.

Now suppose that the representation $T = \sum_{i=0}^{t} d_{2i}u_{2i}$ satisfies (8) and the additional condition. It is not hard to see that then $T$ is maximal for $l = t$, that is, $d_{2t} = aa_{2t+1} + 1$. Now $d_{2k}$ maximal implies $d_{2j} < aa_{2j+1}$ for some $k < j < l$. Since $u_{2i} \geqslant u_{2i-2}$ $(i \geqslant 0)$, $T$ does not decrease if we let $d_{2j} = aa_{2j+1}$ for all $k < j < l$, put $k = 0$ and decrease the maximal value of $d_{2k} = d_0$ by 1. Thus

$$T \leqslant a(a_{2t+1}u_{2t} + a_{2t-1}u_{2t-2} + \cdots + a_3u_2 + a_1u_1)$$
$$+ u_{2t} + \varepsilon a u_0 - 1 = u_{2t+2} + u_{-2} + (\varepsilon a - 1)u_0 - 1 = u_{2t+2} - 1 \qquad (t \geqslant 0)$$

by Lemma 3. Thus condition (1) is satisfied and so the result follows by Theorem 1. ■

In the "Fibonacci case", that is the special case where $\alpha = [\dot{1}]$, the $p_i$-system of Theorem 4 becomes a rather curious ternary numeration system, since (8) now implies $0 \leqslant d_{2i} \leqslant 2$ $(i \geqslant 0)$. In this case the condition of Theorem 4 states that between any two digits 2 there must be a digit 0. This special case has been used by Chung and Graham [**1**] to investigate irregularities of distribution of sequences. It and several variations can also be used for efficient data compression of sparse binary vectors, where the lengths of the 0-runs are expressed in the selected numeration system, and the basis elements are Huffman-coded. (For Huffman coding see [**9**, Sect. 2.3.4.5.].) Numeration systems based on Fibonacci systems of order $m$ and other variations can also be used to design error-insensitive codes for data transmission. The details of these developments will be

elaborated on elsewhere.

Recapping the main results of this article, note that the simple-minded iterated division algorithm is the tool which produces the representation for any nonnegative integer $N$ in any system $S = \{u_0, u_1, u_2, \dots : 1 = u_0 < u_1 < u_2 < \cdots\}$ in every one of the four theorems and the other applications considered here. This representation satisfies (1), and there is only one representation satisfying (1). The various systems considered here differ only in the definition of $S$, precipitating for each case a condition of its own equivalent to (1). This condition is very simple for systems in which the recurrence for the $u_i$ consists of one term only, but may become rather strange-looking for higher order recurrences.

The main upshot of the applications is that—analogously to the case of data structures—the proper choice of a system of numeration may lead to considerably more efficient algorithms.

We finally remark that for representing a negative integer $N$ in any of the above numeration systems, represent $|N|$ and then reverse the signs of all the digits. Note that for representing any integer in this way, its digits are either all nonnegative or all nonpositive. For representations with both positive and negative digits, see Matula [**14**], who considered only the case where the basis elements are powers of some integer $b$, however. Are there any interesting numeration systems using both positive and negative digits, with more general basis elements than $b^i$—such as those considered above?

**References**


**1.** F. R. K. Chung and R. L. Graham, On irregularities of distribution of real sequences, Proc. Nat. Acad. Sci. U.S.A., 78 (1981) 4001.

**2.** S. Even, Algorithmic Combinatorics, Macmillan, New York, 1973.

**3.** A. S. Fraenkel, New proof of the generalized Chinese Remainder Theorem, Proc. Amer. Math. Soc., 14 (1963) 790–791.

**4.** ______, How to beat your Wythoff games' opponent on three fronts, this MONTHLY, 89 (1982) 353–361.

**5.** ______, Wythoff games, continued fractions, cedar trees and Fibonacci searches, Theoret. Comput. Sci., 29 (1984) 49–73.

**6.** A. S. Fraenkel, J. Levitt, and M. Shimshoni, Characterization of the set of values $f(n) = [n\alpha]$, $n = 1, 2, \dots$, Discrete Math., 2 (1972) 335–345.

**7.** A. S. Fraenkel and M. Mor, Combinatorial compression and partitioning of large dictionaries, Comput. J., 26 (1983) 336–343.

**8.** G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 4th edition, 1960.

**9.** D. E. Knuth, The Art of Computer Programming, vol. 1: Fundamental Algorithms, Addison-Wesley, Reading, MA, 1973.

**10.** ______, The Art of Computer Programming, vol. 2: Seminumerical Algorithms, 2nd edition, Addison-Wesley, Reading, MA, 1981.

**11.** ______, The Art of Computer Programming, vol. 3: Sorting and Searching, Addison-Wesley, Reading, MA, 1975.

**12.** D. H. Lehmer, The machine tools of combinatorics, in Applied Combinatorial Mathematics (E. F. Beckenbach, ed.), Wiley, New York, 1964, pp. 5–31.

**13.** W. C. Lynch, The $t$-Fibonacci numbers and polyphase sorting, Fibonacci Quart., 8 (1970) 6–22.

**14.** D. W. Matula, Basic digit sets for radix representation, J. Assoc. Comput. Mach., 29 (1982) 1131–1143.

**15.** M. Mor and A. S. Fraenkel, Cayley permutations, Discrete Math., 48 (1984) 101–112.

**16.** C. D. Olds, Continued Fractions, New Mathematical Library, vol. 9, Mathematical Association of America, Washington, DC, 1963.

**17.** O. Perron, Die Lehre von den Kettenbrüchen, Band I, Teubner, Stuttgart, 1954.

**18.** M. J. Whinihan, Fibonacci Nim, Fibonacci Quart., 1 (No. 4, 1963) 9–12.

**19.** A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, translated by J. McCawley, Jr., revised and edited by B. Gordon, vol. II, Holden-Day, San Francisco, 1967.

**20.** E. Zeckendorf, Représentation des nombres naturels par une somme de nombres de Fibonacci ou de nombres de Lucas, Bull. Soc. Roy. Sci. Liège, 41 (1972) 179–182.

# AN ELEMENTARY PROOF OF THE CAUCHY-KOWALEVSKY THEOREM

WOLFGANG WALTER
*Universität Karlsruhe, Mathematisches Institut I, Kaiserstr. 12, D-7500 Karlsruhe, West Germany*

**Dedicated to Professor Johannes Weissinger on his seventieth birthday***

**0. Introduction.** It is the main aim of this article to present an elementary proof of the Cauchy-Kowalevsky (= C − K) theorem based on the contraction principle. The theorem deals with a nonlinear partial differential equation

$$D_t^p u = F\big(t, x_1, \dots, x_n, u, \dots, D_t^k D_x^\alpha u, \dots\big) \qquad (k + |\alpha| \leqslant p;\ k < p) \tag{1}$$

or with a system of such equations, where $D_t = \partial/\partial t$, $D_x^\alpha = \partial^{|\alpha|}/\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}$, and $p \geqslant 1$. It states that if for $t = 0$ initial data $D_t^k u = \phi_k$ $(0 \leqslant k < p)$ are prescribed, where the $\phi_k$ are analytic functions of $x = (x_1, \dots, x_n)$ in a neighborhood of $x_0 = (x_1^0, \dots, x_n^0)$, and if the function $F$ is analytic in a neighborhood of the point $(0, x_0, \dots, D_x^\alpha \phi_k(0, x_0), \dots)$, then there exists a unique solution $u(t, x)$ in a neighborhood of $(0, x_0)$, analytic in $t$ and $x$ and satisfying the initial conditions. The problem is called *Cauchy's problem*, the initial conditions *Cauchy data*. It was first solved by Cauchy [**2**], and, in a more general and simplified way, by Sophie v. Kowalevsky [**9**]. It represents the analogue for partial differential equations of the basic initial value problem for an ordinary differential equation $u^{(p)}(t) = f(t, u, u', \dots, u^{(p-1)})$ with given initial data $u(0), \dots, u^{(p-1)}(0)$. Obviously, this analogy motivated Cauchy to consider the problem.

As a preliminary step of the proof found, e.g., in [**3**; p. 39-56], the equation (or system) is first converted into a more convenient quasilinear first order system

$$u_t = \sum_{j=1}^{n} B_j(t, x, u)\, u_{x_j} + c(t, x, u). \tag{2}$$

In the vector notation used here, $x = (x_1, \dots, x_n)$, $u = (u_1, \dots, u_m)$, the $B_j$ are $m$ by $m$ matrices, and $u$, $c$ are 1 by $m$ matrices, i.e., column vectors. Some textbooks, e.g. [**10**] [**18**], consider only the linear case

$$u_t = A(t, x) u + \sum_{j=1}^{n} B_j(t, x)\, u_{x_j} + c(t, x), \tag{3}$$

which is much easier to handle. In both cases the initial condition is given by $u(0, x) = \phi(x)$.

We show in the important special case of a nonlinear second order problem (with $n = 1$, for simplicity)

$$u_{tt} = f(t, x, u, u_t, u_x, u_{tx}, u_{xx}),\ u(0, x) = \phi_0(x),\ u_t(0, x) = \phi_1(x), \tag{1'}$$

---

*Wolfgang Walter*: I was born in Schwäbisch Gmünd/Germany. Since my early age, mathematics and music have been my main interests. Keeping the latter field for leisure, I studied mathematics and physics at the Universität Tübingen and received a Doctorate in 1956 under E. Kamke. In 1958 I followed an invitation by A. Weinstein and spent a year at the University of Maryland. Since 1963 I am Professor of Mathematics at the Universität Karlsruhe, spending a considerable part of my professional life at various American universities. My main interest is in ordinary and partial differential equations with a special emphasis in differential inequalities. I have written the monograph *Differential and Integral Inequalities* (Ergebnisse der Mathematik, Vol. 55) and several mathematical textbooks. The main ideas of the present article emerged in 1982/83 while I was a Visiting Professor at the University of Florida and the Georgia Institute of Technology.

*As early as 1952 Professor Weissinger suggested to use contraction proofs in the classroom. His article *Zur Theorie und Anwendung des Iterationsverfahrens* (Math. Nachr., 8, 193–212, 1952) contains the generalized contraction principle and numerous applications from different branches of pure and applied analysis.

how this transformation is accomplished. Setting $u_t = v$ and $u_x = w$, a first order problem

$$(1'') \quad \begin{cases} u_t = v, & u(0,x) = \phi_0(x), \\ v_t = f(t,x,u,v,w,v_x,w_x), & v(0,x) = \phi_1(x), \\ w_t = v_x, & w(0,x) = \phi_0'(x), \end{cases}$$

is obtained. It is easily seen that $u$ solves (1′) if and only if $(u, v, w)$ solves (1″) (passing from the latter equation to the former one requires the equation $w = u_x$, which follows from $u_{tx} = w_t$ by integration in the $t$ direction). Now, (1″) or, more generally, any first order system of the form

$$u_t = g(t,x,u,u_x) \text{ with } u = (u_1,\dots,u_m), \qquad g = (g_1,\dots,g_m)$$

transforms by differentiation with respect to $x$ into a quasilinear system for $(u,v) = (u,u_x)$

$$(2') \quad \begin{cases} u_t = g(t,x,u,v), \\ v_t = g_x + g_u v + g_v v_x, \end{cases}$$

the argument being $(t,x,u,v)$ in all three terms. The reasoning above is easily adapted to the case where $x = (x_1,\dots,x_m)$ and/or $u = (u_1,\dots,u_m)$ in (1′).

In the classical method of proof, used by Cauchy **[2]** and Sophie v. Kowalevsky[1] **[9]** and presented in many textbooks, the solution is sought as a power series in $t$ and $x_i$ whose coefficients are determined by recursion formulas using the corresponding expansions of $B_j$ and $c$. The centerpiece of the approach is the convergence proof for this series by the *method of majorants*. It consists of considering an auxiliary problem of the same structure (with the same recursion formulas) which majorizes the given problem in the sense that all coefficients involved are positive and larger than the absolute values of those in the given problem. Cauchy invented and applied the method in many instances. He called it, not very suggestively, *calcul des limites*.

A new proof based on entirely different ideas was discovered in 1941 by M. Nagumo **[11]**. Nagumo considers the quasilinear system (2) with zero initial values (no loss of generality) and transforms it into an equivalent equation for $v = u_t$

$$v(t,z) = \sum_{j=1}^{n} B_j(t,z,Iv)\, Iv_{z_j} + c(t,z,Iv),$$

where $(Iv)(t,z) = \int_0^t v(\tau,z)\,d\tau \; (= u)$. This equation is treated as an operator equation $v = \Phi(v)$ in an appropriate Banach space $E$, and a solution is obtained as a fixed point of $\Phi$, using the Schauder fixed point theorem. An essential tool is a lemma, reproduced with proof in Section 2, which provides an estimate of the derivative of a holomorphic function in terms of the function itself (in **[11]** it is given for the case where $\Omega$ is a ball $|z| < R$). This lemma has proved to be of importance in many modern developments of the Cauchy problem and is called here *Nagumo's lemma*. As a consequence, the variables $x_i$ are now complex by necessity, and for this reason we have written $z$ instead of $x$. On the other hand, the variable $t$ is assumed to be real, and the solution is shown to exist in a conical region $0 \leqslant t < l$, $|z| < R - Lt$. The functions $B_j$ and $c$ are assumed to be continuous in all variables and holomorphic in $z$ and $u$, and $v$ is continuous in $(t,z)$ and holomorphic in $z$. Strictly speaking, Nagumo's theorem does not cover the C-K theorem. But its proof is also valid in the case where $t$ is a complex variable and the functions involved are holomorphic in all variables, a fact which Nagumo does not mention. In short, Nagumo obtained a new proof and, by allowing $t$ to be a real variable, an important generalization of the C-K theorem. But this approach has also a drawback: uniqueness does not follow from the Schauder fixed point theorem and has to be proved separately. Another point of comparison

[1] Different ways of spelling the name have appeared in the more recent literature (Kowalewska, Kowalewskaya, ...). The author, who has neither the historic nor the linguistic expertise to make a reasonable choice, has been led by the doctoral dissertation **[9]** which was written by "Frau Sophie von Kowalevsky".

regards the "real" version of the C-K theorem, which states that the solution is real-valued for real values of the variables if the coefficients in the differential equation have this property. Nagumo does not prove that fact, yet it is easily obtained in the framework of his proof. If one restricts either the whole Banach space or the convex, compact subset which is mapped into itself by $\Phi$, to functions which are real-valued for real $z_j$, then his proof gives the desired result. Recently, Keller and Schneider [**8**] published a proof of the classical C-K theorem by reduction to Schauder's fixed point theorem, which bears a likeness to Nagumo's proof (the latter is not cited).

Another approach which gives rise to various generalized versions of the C-K theorem was developed in the sixties. The basic notion is that of a *scale of Banach spaces*, that is a collection $(B_\rho)_{\rho>0}$ of Banach spaces with the property that $0<\rho<\sigma$ implies $B_\sigma \subset B_\rho$ and $\|u\|_\rho \leqslant \|u\|_\sigma$ for $u \in B_\sigma$. One considers bounded linear operators $L$ from $B_\sigma$ into $B_\rho$, $\|Lu\|_\rho \leqslant \|L\|_\rho^\sigma \|u\|_\sigma$ for any $0<\rho<\sigma$, which satisfy an estimate

$$\|L\|_\rho^\sigma \leqslant \frac{c}{\sigma-\rho}.$$

These notions are already used in Vol. III of Gelfand and Silov's treatise [**4**], A. 2.1-2.3, in connection with the linear Cauchy problem for constant coefficients and analytic initial values. In 1965, Ovsjannikov [**14**] proved an existence and uniqueness theorem, often quoted as *Ovsjannikov's theorem*, for the Cauchy problem

$$\frac{du}{dt} = L_t u, \qquad u(0) = u_0 \in B_\rho$$

in this abstract framework, where $(L_t)$ is a family of linear operators with the properties stated above for $L$. Further research along these lines by Treves [**20**] [**21**], Nirenberg [**12**], Ovsjannikov [**15**] [**16**], Nishida [**13**], Pate [**17**] and others (see also the bibliography of these articles for further reading) led to linear and nonlinear "abstract" Cauchy-Kowalevsky theorems. The connection with the classical theorem is established by the example

$$B_\rho = \left\{ u\colon u(z) \text{ is holomorphic for } \max|z_i| < \rho,\ \|u\|_\rho = \sup_{|z|<\rho} |u(z)| < \infty \right\}$$

and the operators $L = \partial/\partial z_i\colon B_\sigma \to B_\rho$ which satisfy $\|\partial/\partial z_i\|_\rho^\sigma \leqslant 1/(\sigma-\rho)$ for $0<\rho<\sigma$. An easily accessible exposition of the linear case is given in Chapter 17 of Treves' book [**22**]. This theory shares with Nagumo's proof the property that $t$ can be a real or complex variable.

The proof presented in this paper is elementary inasmuch as it uses Banach's contraction principle as the only tool from functional analysis. Basic properties of holomorphic functions $u(\zeta)$ depending on a single complex variable $\zeta$ such as Cauchy's integral formula are used. As far as holomorphic functions of several complex variables are concerned, it is sufficient to know the definition (a function $f\colon G \to \mathbb{C}$, where $G$ is an open subset of $\mathbb{C}^n$, is holomorphic in $G$, if $f$ and the derivatives $\partial f/\partial z_i$ $(i = 1,\dots,n)$ are continuous in $G$) and the fact that $f(u_1(\zeta),\dots,u_n(\zeta))$ is holomorphic in $\zeta$, if $f$ and the $u_i$ are holomorphic. Power series expansions, which can be used to give an equivalent definition of holomorphy, are not used. The exposition is written with an eye to classroom use. For this reason, the linear case is treated first, where the proof is remarkably simple and short. Apart from didactic advantages, the proof has an important mathematical consequence: it gives not only existence and uniqueness, but also estimates between the solution and an approximate solution, in particular, continuous dependence on the initial values and the right hand side of the differential equation. In short: the proof by contraction shows that the analytic Cauchy problem is a well-posed (or correctly set) problem.

At first glance this statement seems to contradict well-known facts about second order partial differential equations of mathematical physics. In explaining the situation, we consider the simplest prototypes, (a) the hyperbolic (vibrating string) equation $u_{tt} = u_{xx}$, (b) the elliptic (potential) equation $u_{tt} + u_{xx} = 0$, and (c) the parabolic (heat) equation $u_t = u_{xx}$. In the hyperbolic case, the problem models a simple physical situation. It is correctly set, and it has a

solution not only for analytic, but also for much more general (e.g., continuous) initial values. In contrast, the classical, well-posed elliptic problems are not Cauchy problems, but boundary value problems where only one function (e.g., $u$) is prescribed on the whole boundary. Consider, for example, problem (b′) with initial values $u(0,x) = 0$, $u_t(0,x) = \phi_1(x)$. If a solution exists, it is a harmonic and hence (real) analytic function for $t > 0$. Furthermore, the solution allows an analytic continuation for $t \leqslant 0$ given by $u(-t,x) = -u(t,x)$. Consequently, $\phi_1$ is analytic. Thus problem $(b')$, while possessing a solution for analytic $\phi_1$, by virtue of the C-K theorem, never has a solution for non-analytic $\phi_1$. A more detailed account is given in Hadamard's classical *Lectures on Cauchy's problem* [**5**, Chapter II]. Here one finds also the famous example (p. 33):

$$u_{tt} + u_{xx} = 0, \quad u(0,x) = 0, \quad u_t(0,x) = a \sin kx,$$

with the solution $u(t,x) = (a/k)\,(\sin kx)(\sinh kt)$. If $a > 0$ is small, then $|u(0,x)| \leqslant a$ is small, but for any fixed $t \neq 0$ the solution can be made arbitrarily large by choosing $k$ large. Hence the zero solution $(a = 0)$ does not depend continuously on the initial values. These considerations give the impression that the applicability of the C-K theorem to elliptic problems is a strange curiosity: the solution exists, but it behaves oddly.

But the odd behavior appears only as long as we consider analytic functions for *real* variables (one might add that there are other cases of odd behavior in this setting). If we put holomorphic functions in their natural habitat, which is $\mathbb{C}$ or $\mathbb{C}^n$, the oddities vanish. Theorem 2 will make this point more precise. It is followed by a discussion of Hadamard's example.

The heat equation of problem (c) is not of the type of equation (1); in this equation we have $p = 1$, but $|\alpha| = 2$ in violation of the condition $|\alpha| \leqslant p$. The C-K theorem becomes false for the heat equation. The following counterexample is already found in [**9**, p. 22]. The problem $u_t = u_{xx}$, $u(0,x) = 1/(1-x)$ has a unique formal power series solution

$$u(t,x) = \sum_{i=0}^{\infty} \frac{(2i)!\,t^i}{i!(1-x)^{2i+1}} = \sum_{i,k=0}^{\infty} \frac{(2i)!}{i!k!} t^i x^k,$$

but the series diverges for any $t \neq 0$.

In this connection we remark that well-posedness, while being a standard subject with regard to initial and boundary value problems for elliptic, hyperbolic,... partial differential equations, is only rarely discussed in the Cauchy problem with analytic data. Also, the notion of well-posedness is often defined in a narrow sense, meaning continuous dependence on initial values only, while the physical meaning of this notion requires that continuous dependence on the right hand side of the differential equation be included.

Let us finally note that there is a more general Cauchy problem for equation (1), where Cauchy data (i.e., initial values) are given on a noncharacteristic hypersurface $\Omega$ defined by an equation $\phi(t, z_1, \ldots, z_n) = 0$. This case can be reduced—at least locally—to the case treated here where $\Omega$ is lying in the hyperplane $t = 0$. The notion of noncharacteristic surface and the reduction procedure are described in many textbooks, e.g., in [7; Chapter 3] and [**18**; § 3].

The author has looked for a proof of the C-K theorem by contraction for some time. His conviction that such a proof must exist was based on Bessaga's noteworthy theorem [**1**]:

*If $M$ is a set, $\alpha \in (0,1)$ and $T$: $M \to M$ a map with the property that $T^n$ has one and only one fixed point for $n = 1, 2, \ldots,$ then there exists a metric $d$ on $M$ which makes $M$ a complete metric space and for which $d(Tx, Ty) \leqslant \alpha d(x,y)$.*

The Cauchy problem in question can be written as a fixed point equation $u = Tu$ (see (9) for the linear case and (18) for the nonlinear case), and the C-K theorem states that $T$ has a unique fixed point. Moreover, it is not difficult to show that $T^n$ also has a unique fixed point for any $n \geqslant 1$. Bessaga's theorem now implies that $T$ is a contraction in a suitable metric space. The problem was to find the metric.

**1. Geometrical preliminaries.** Since we want to construct a solution in the large, some

geometrical considerations are necessary. Let $\Omega$ be an open set in $\mathbb{C}^n$ with a nonempty boundary $\Gamma = \partial\Omega$, and let $d(z) = \operatorname{dist}(z, \Gamma)$ be the distance from $z \in \Omega$ to $\Gamma$, measured in the maximum norm $|z| = \max_{i=1,\dots,n}|z_i|$. The set $G$, a subset of $\mathbb{R} \times \mathbb{C}^n$ (real case) or $\mathbb{C}^{n+1}$ (complex case), consists of all points $(t, z)$ with $z \in \Omega$ and $|t| < \eta d(z)$, where $\eta > 0$ will be specified below. Let $\Omega_t$ be the set of all $z \in \Omega$ such that $(t, z) \in G$ or, equivalently, $d(z) > |t|/\eta$. It is easily seen that

$$d(t,z) = d(z) - \frac{|t|}{\eta} > 0 \qquad (z \in \Omega_t) \tag{4}$$

is the distance from $z$ to the boundary $\Gamma_t = \partial\Omega_t$. In geometrical terms, $G$ is the double cone with base $\Omega$ and slope $\eta$, and $\Omega_t$ is the base of that part of $G$ which lies above $t$ ($t > 0$) or below $t$ ($t < 0$). This description applies to the real case. The following property of $d(t,z)$ will be needed later:

$$z \in \Omega_t, |z - z'| = r < d(t,z) \Rightarrow z' \in \Omega_t \text{ and } d(t,z') \geqslant d(t,z) - r. \tag{5}$$

This follows readily from the triangle inequality $d(z) \leqslant |z - z'| + d(z')$.

If one is interested only in a local solution, it suffices to consider the case where $\Omega$ equals $B_R(z_0)$, the open ball in $\mathbb{C}^n$ with center at $z_0$ and radius $R$. In this case $d(z) = R - |z - z_0|$, $\Omega_t = B_r(z_0)$ with $r = R - |t|/\eta$ and $d(t,z) = R - |z - z_0| - |t|/\eta$.

**2. The linear case.** The linear C-K theorem deals with the following initial value problem

$$u_t = A(t,z)u + \sum_{j=1}^{n} B_j(t,z)u_{z_j} + c(t,z) \text{ in } G, \tag{6}$$

$$u(0,z) = \phi(z) \text{ in } \Omega, \tag{7}$$

which can be stated as an equivalent integral equation of Volterra type

$$u(t,z) = g(t,z) + \int_0^t \left[ A(\tau,z)u(\tau,z) + \sum_{j=1}^{n} B_j(\tau,z)u_{z_j}(\tau,z) \right] d\tau, \tag{8}$$

where

$$g(t,z) = \phi(z) + \int_0^t c(\tau,z)\, d\tau.$$

Here, $u = (u_1, \dots, u_m)$ has values in $\mathbb{C}^m$, $A$ and $B_j$ are complex-valued $m$ by $m$ matrices, and $c$ and $f$ are complex-valued $m$-vectors. In our matrix notation, $u$, $u_t = \partial u/\partial t$, $u_{z_j} = \partial u/\partial z_j$, $\phi$ and $c$ are considered column vectors. A solution of (8) is, by definition, continuous in $G$ and holomorphic in $z$ for fixed $t$ (real case) or holomorphic in $t$ and $z$ (complex case). In the complex case it is understood that integration is performed along the straight line from 0 to $t$. In both cases, a solution of (8) is, under proper assumptions regarding $A, \dots, \phi$, of class $\mathbb{C}$ in $G$ (complex derivatives with respect to the complex variables) and a solution of the initial value problem (6) (7), and vice versa.

To avoid ambiguity, we use the maximum norm $|u| = \max|u_j|$ $(1 \leqslant j \leqslant m)$ and the corresponding operator norm for matrices $A = (a_{jk})$, $|A| = \max_k \Sigma_j |a_{jk}|$, but any other pair of norms with the property $|Au| \leqslant |A||u|$ would do as well.

The following lemma, which goes back to Nagumo [**11**], gives a bound for $\partial f/\partial z_j$ in terms of $f$ and is essential for the proof.

NAGUMO'S LEMMA. *Let $f\colon \Omega \to \mathbb{C}^m$ be holomorphic and $p \geqslant 0$. Then*

$$|f(z)| \leqslant \frac{c}{d^p(z)} \Rightarrow |f_{z_j}(z)| \leqslant C_p \frac{c}{d^{p+1}(z)},$$

*where* $C_p = (1 + p)\left(1 + \dfrac{1}{p}\right)^p < e(p+1)$, $C_0 = 1$.

*Proof.* For a function $\psi$ of a single complex variable $\zeta$, which is holomorphic in the circle $|\zeta - \zeta'| \leqslant r$, the estimate

$$|\psi'(\zeta)| \leqslant \frac{1}{r} \max_{|\zeta-\zeta'|=r} |\psi(\zeta')|$$

holds. It follows immediately from the Cauchy integral representation formula

$$\psi'(\zeta) = \int \psi(\zeta')/\left[2\pi i(\zeta' - \zeta)^2\right] d\zeta',$$

in which the integral extends over the circle $|\zeta - \zeta'| = r$. Applying this result with $\zeta = z_j$, we obtain the inequalities

$$|f_{z_j}(z)| \leqslant \frac{1}{r} \max_{|z-z'|=r} |f(z')| \leqslant \frac{c}{r} \max \frac{1}{d^p(z')} \leqslant \frac{c}{r(d-r)^p}.$$

Here, $0 < r < d \equiv d(z)$ and $d(z') \geqslant d - r$ because of (2). The choice $r = d/(p+1)$ gives the (optimal) value stated in the lemma. □

To be specific, the following theorem is formulated for the real case.

THEOREM 1. *Assume that*

(i) *the functions* $A(t,z)$, $B_j(t,z)$, $c(t,z)$ *are continuous in* $G$ *and holomorphic in* $z$ *for fixed* $t$, *the function* $f(z)$ *is holomorphic in* $z$;

(ii) *there exist positive constants* $\alpha$, $\beta_j$, $\gamma$, $\delta$ *and* $p$ *such that*

$$|A(t,z)| \leqslant \frac{\alpha}{d(t,z)}, \qquad |B_j(t,z)| \leqslant \beta_j,$$

$$|c(t,z)| \leqslant \frac{\gamma}{d^{p+1}(t,z)}, \qquad |f(z)| \leqslant \frac{\delta}{d^p(z)} \text{ in } G.$$

(iii) $\alpha/p + (1 + (1/p))^{p+1}\Sigma\beta_j < 1/\eta$ (*this can always be achieved by diminishing* $\eta$, *if necessary*).

*Then Equation* (8) *has a unique solution* $u$ *in* $G$, *and it satisfies* $|u(t,z)| \leqslant C/d^p(t,z)$ *in* $G$.

*Proof.* Let $E$ be the Banach space of all functions $u \in C^0(G, \mathbb{C}^m)$, which are holomorphic in $z$ and have a finite norm

$$\|u\| := \sup_G |u(t,z)| d^p(t,z).$$

Note that convergence in the norm implies uniform convergence in compact subsets of $G$; hence the limit is holomorphic in $z$ and $E$ is complete. We write Equation (8) in the form

$$u = g + Tu,$$

where $T$ is a linear operator given by

$$(Tu)(t,z) = \int_0^t \left[ A(\tau,z)u(\tau,z) + \sum_j B_j(\tau,z)u_{z_j}(\tau,z) \right] d\tau. \tag{9}$$

It follows easily from the inequality

$$\left| \int_0^t \frac{d\tau}{d^{p+1}(\tau,z)} \right| = \int_0^{|t|} \frac{ds}{(d(z) - s/\eta)^{p+1}} < \frac{\eta}{pd^p(t,z)} \tag{10}$$

that $g \in E$. According to the definition of the norm, we have

$$|u(t,z)| \leqslant \frac{\|u\|}{d^p(t,x)}. \tag{11}$$

Nagumo's lemma, applied to the region $\Omega_t$ with distance function $d(t,z)$ instead of $\Omega$ and $d(z)$, gives the esimtate

$$|u_{z_j}(t,z)| \leqslant C_p \frac{\|u\|}{d^{p+1}(t,z)}. \tag{12}$$

It follows from the assumption (ii) together with (11) and (12) that

$$|Au| \leqslant \frac{\alpha\|u\|}{d^{p+1}(t,z)}, \; |B_j u_{z_j}| \leqslant \frac{\|u\|}{d^{p+1}(t,z)}\beta_j C_p,$$

hence, with $\beta = \Sigma\beta_j$,

$$\begin{aligned} |(Tu)(t,z)| &\leqslant \|u\|(\alpha + \beta C_p)\left|\int_0^t \frac{d\tau}{d^{p+1}(\tau,z)}\right| \\ &\leqslant \frac{1}{p}(\alpha + \beta C_p)\eta\frac{\|u\|}{d^p(t,z)}. \end{aligned}$$

In the last step, inequality (10) was used. Going back to the definition of the norm, we arrive at the final estimate

$$\|Tu\| \leqslant q\|u\|, \qquad \text{where } q = \left(\frac{\alpha}{p} + \beta\left(1 + \frac{1}{p}\right)^{p+1}\right)\eta < 1, \tag{13}$$

according to assertion (iii). Hence the contraction principle applies to the equation $u = g + Tu$, and it follows that this equation has exactly one solution in $E$. This proves the theorem except for the possibility (which will be excluded below) that there are other solutions which do not belong to $E$. □

REMARK 1. It is seen from (iii) that the constant $\eta$, which describes the region of existence of the solution, depends only on $\beta = \Sigma_{j=1}^n \beta_j$. Indeed, for $p$ large, $\alpha/p$ is small and $\left(1 + \frac{1}{p}\right)^{p+1}$ is close to $e$ (note that $\alpha$ and $\beta$ do not depend on $p$). Hence (iii) is satisfied (for large $p$) if $\eta < 1/(e\beta)$. Very simple examples show that this result is, except for the factor $e$, optimal. Let $u_t = bu_z$ and $u(0,z) = \phi(z)$ ($n = 1$, $b$ constant). Then the solution is given by $u(t,z) = \phi(bt + z)$. If $\phi$ is holomorphic, say, in the circle $|z| < 1$, then the solution exists for $|z + bt| < 1$, and this is best possible when $\phi$ cannot be continued analytically beyond the unit circle. If we vary $b$, keeping $|b| = \beta$ fixed, then the largest region common to all those regions of existence is the circular cone $\beta|t| < 1 - |z|$, i.e., $\eta = 1/\beta$ is best possible.

REMARK 2. In order to understand the nature of the assumptions (ii), let us assume that $n = 1$ and that $z_0$ is an isolated boundary point of $\Omega$, i.e., that $z \in \Omega$ and $d(z) = |z - z_0|$ for $z$ near $z_0$. Then $A$ is allowed to have a pole of the first order, while $f$ and $c$ may have a pole of any order at $z_0$. Now let $\Omega$ be unbounded (this case is usually not considered in the context of the C-K theorem). Then $d(t,z) \to \infty$ (perhaps) as $|z| \to \infty$, which implies that $A$, $c$ and $\phi$ must vanish at infinity. If, on the other hand, assumption (ii) is satisfied, then the region of existence $G$ extends to infinity in the $t$ direction. By a slight change in the proof one can handle the case where $A$, $c$ and $\phi$ do not vanish at infinity at the price of a smaller region of existence which is bounded in $t$. The change concerns the definition of $d(z)$. Instead of the distance to $\Gamma$ we may take $d(z) = \min(\text{dist}(z,\Gamma), d_0)$, where $d_0$ is a positive constant, or more generally, any function with the properties

$$0 < d(z) \leqslant \text{dist}(z,\Gamma) \quad \text{and} \quad |d(z) - d(z')| \leqslant |z - z'| \quad \text{for } z, z' \in \Omega. \tag{14}$$

If $G$, $\Omega_t$ and $d(t,z)$ are defined as before, using the new function $d(z)$, then property (5) and its one-line proof remain valid, and $0 < d(t,z) \leqslant \text{dist}(z,\Gamma_t)$ for $z \in \Omega_t$. As a consequence, the

lemma remains true, and the proof of the theorem goes through. If we choose a bounded function $d(z)$, then the inequalities in assumption (ii) can be satisfied as long as $A$, $c$ and $f$ remain bounded for $|z| \to \infty$.

REMARK 3. In the complex case it is assumed that the functions $A$, $B_j$ and $c$ are holomorphic in $t$ and $z$. Otherwise the assumptions are unchanged. Theorem 1 remains true, and the proof goes through with the obvious change that $E$ is now the Banach space of all functions $u$ which are holomorphic in $t$ and $z$ and have a finite norm. Note that (10) remains true as it stands.

REMARK 4. In order to exclude the possibility that there are other solutions not belonging to $E$, let us assume that $u^*$ is a solution to (6) (7) in some open domain $D \subset G$. Let $(0, z_0) \in D$, let $\Omega^* = B_r(z_0)$ be a ball in $\mathbb{C}^n$ with small radius $r$, and let $G^*$ be the corresponding set in $\mathbb{R} \times \mathbb{C}$ defined by the inequalities $|t| < \eta d^*(z)$, where now $d^*(z) = r - |z - z_0|$. Applying our theorem to the set $\Omega^*$ we see that there is exactly one solution in $G^*$ belonging to the corresponding Banach space $E^*$. But since our original solution $u$ as well as the solution $u^*$ (strictly speaking, their restrictions to $G^*$) belong to $E^*$, we arrive at $u = u^*$ in $G^*$. This argument, together with the fact that solutions are holomorphic in $z$, leads to $u = u^*$ at least in a small strip, say, for $(t, z) \in D$ and $|t| \leqslant \alpha$. But now the reasoning can be repeated with initial values given for $t = \pm\alpha$ instead of $t = 0$. In this way one obtains $u = u^*$ in $D$. In the complex case the last step is not necessary, since $u = u^*$ in the small implies equality in the large.

REMARK 5. Assume that the functions in Theorem 1 have the property that their components are real-valued for real values of $z_j$ and $t$. Then the solution $u$ has the same property. This well-known fact follows immediately from our proof. Indeed, the function $g(t, z) = \phi(z) + \int_0^t c(s, z)\, ds$ has the property in question. If $v$ has the property, then also $Tv$. Hence, all members of the sequence $(u_k)$ defined by $u_0 = g$, $u_{k+1} = g + Tu_k$ $(k = 0, 1, \ldots)$ have the property, and so does the solution $u = \lim u_k$.

The fixed point of a contractive operator $S$ is not only unique; it also depends continuously on $S$. Our next theorem deals with this important consequence of the contraction principle.

THEOREM 2. *Under the assumptions of Theorem* 1, *the solution $u$ of the Cauchy problem* (6) (7) *depends continuously on the given functions $A$, $B_j$, $c$ and $f$: For any given $\varepsilon > 0$, there exists $\delta > 0$ such that the solution $u^*$ of a corresponding problem* (6) (7) *with $A$, $B_j$, $c$ and $\phi$ replaced by corresponding starred quantities which satisfy the inequalities*

$$|A(t,z) - A^*(t,z)| \leqslant \frac{\delta}{d(t,z)}, \qquad |B_j(t,z) - B_j^*(t,z)| \leqslant \delta,$$

$$|c(t,z) - c^*(t,z)| \leqslant \frac{\delta}{d^{p+1}(t,z)}, \qquad |\phi(z) - \phi^*(z)| \leqslant \frac{\delta}{d^p(z)},$$

belongs to the $\varepsilon$-neighborhood of $u$ (in $E$),

$$|u(t,z) - u^*(t,z)| \leqslant \frac{\varepsilon}{d^p(t,z)}.$$

*Proof*. The assertion follows easily from a standard estimate for contractive mappings. Let $S$ be any contraction, $\|Su - Sv\| \leqslant q\|u - v\|$, and let $u$ be the fixed point of $S$. Then

$$\|u - v\| \leqslant \|Su - Sv\| + \|Sv - v\| \leqslant q\|u - v\| + \|Sv - v\|,$$

which implies

$$\|u - v\| \leqslant \frac{1}{1-q}\|v - Sv\|. \tag{15}$$

As in the proof of Theorem 1, we write $u = g + Tu$ and analogously $u^* = g^* + T^*u^*$. Then (15) reads

$$\|u - u^*\| \leqslant \frac{1}{1-q}\{\|g - g^*\| + \|(T - T^*)u^*\|\}. \tag{16}$$

It follows easily from our assumptions and (10) that $\|g - g^*\| \leqslant \delta C$, $C = 1 + \eta/p$. The estimate (13), applied to $T - T^*$ (which means $\alpha = \beta = \delta$) shows that $\|T - T^*\| \leqslant \delta C_1$. It also follows from (13) that for $0 \leqslant \delta < \delta_0$, where $\delta_0$ is small, $\|T^*\| \leqslant q^* < 1$ with $q^*$ independent of $\delta$. Hence $u^*$ is bounded independent of $\delta$ (take, for example, (15) with $u$, $v$, $q$ and $S$ replaced by $u^*$, 0, $q^*$ and $g^* + T^*$). Summing up, we have obtained an estimate $\|u - u^*\| \leqslant \delta C_2$ with $C_2$ independent of $\delta$, which proves the theorem. □

REMARK. We come back to Hadamard's example ($a = 1$)

$$u_{tt} + u_{xx} = 0, \qquad u(0,x) = 0, \qquad u_t(0,x) = \sin kx,$$

or, in the form of a first order system ($v = u_t, w = u_x$),

$$u_t = v, \qquad v_t = -w_x, \qquad w_t = v_x,$$
$$u(0,x) = w(0,x) = 0, \qquad v(0,x) = \sin kx.$$

We replace $x$ by $z$, choose the ball $|z| < \rho$ for $\Omega$, and consider this example with regard to Theorems 1 and 2. It is easily seen that in Theorem 1 we have $|A| = 1$, $|B| = 1$, $\alpha = \beta = 1$; hence $G$ is defined by $|t| < \eta(\rho - |z|)$ with $\eta < 1/e$. The solution is given by

$$u(t,z) = \frac{1}{k} \cdot \sin kz \cdot \sinh kt.$$

The maximum of $|\sin kz|$ or $|\cos kz|$ for $|z| \leqslant \sigma$ is approximately $\frac{1}{2}e^{k\sigma}$, the maximum of $|\sinh kt|$ or $|\cosh kt|$ for $|t| \leqslant \tau$ is approximately $\frac{1}{2}e^{k\tau}$, and in $G$ we have $|t| + \eta|z| < \rho$, which implies $|t| + |z| < \rho$. Hence for $k \to \infty$ the solution $(u, u_t, u_x)$ does not grow faster in $G$ than on $\Omega$.

**3. The quasilinear case.** As pointed out in the introduction, the case of a general nonlinear system reduces to that of a quasilinear system of the first order

$$u_t = \sum_{j=1}^{n} B_j(t,z,u)u_{z_j} + c(t,z,u). \tag{17}$$

As before, the $B_j$ are $m$ by $m$ matrices, and $c$ is a column vector of order $m$. This system is dealt with in our next theorem. As in the linear case, we describe a region of ascertained existence for the solution in terms of the given functions. But it is not our aim to give an optimal region, which would be somewhat cumbersome. The domain $G = G_\eta$ is defined as before by the inquality $|t| < \eta d(z)$, and $B_R$ is the open ball $|z| < R$ in $\mathbf{C}^m$. In order to avoid complications as described in Remark 2, it is assumed from the beginning that the "distance function" $d(z)$ is bounded and, of course, satisfies (14) (for example, $\Omega$ bounded). If the initial condition is given by $u(0,z) = \phi(z)$ in $\Omega$, we write the solution in the form $u = \phi + v$. This gives a new differential equation for $v$, which is of the same type as (17), and reduces the initial values to zero. We shall therefore restrict consideration to the latter case.

THEOREM 3. *Assume that the functions $B_j(t,z,u)$ and $c(t,z,u)$ are continuous in $G_\eta \times B_R$ and holomorphic with respect to $z$ and $u$ (real case), or holomorphic in $G_\eta \times B_R$ with respect to all variables (complex case), and that the following estimates hold there*:

$$|c| \leqslant \frac{\gamma}{\sqrt{d(t,z)}}, \qquad d(t,z)|c(t,z,u) - c(t,z,v)| \leqslant \gamma'|u - v|,$$

$$|B_j| \leqslant \beta_j, \qquad \sqrt{d(t,z)}\,|B_j(t,z,u) - B_j(t,z,v)| \leqslant \beta_j'|u - v|$$

*($j = 1,\dots,n$). If $\eta > 0$ is such that*

$$2\eta\sqrt{d_0}(\beta + \gamma) < R, \qquad \eta(3\sqrt{3}(\beta + \gamma) + 2\beta) \leqslant 1 \quad \textit{and} \quad \eta e\beta' < 1$$

*with* $d_0 = \sup d(z) < \infty$, $\beta = \Sigma\beta_j$ *and* $\beta' = \Sigma\beta_j'$, *then the quasilinear system* (17) *has a unique solution* $u$ *satisfying the initial condition* $u(0,z) = 0$ *for* $z \in \Omega$ *and existing in* $G_\eta$.

*Proof*. As in the linear case we consider the corresponding integral equation

$$u(t,z) = \int_0^t \left[ \sum_j B_j(\tau, z, u) u_{z_j}(\tau, z) + c(\tau, z, u) \right] d\tau \tag{18}$$

and treat it as a fixed point equation $u = Su$ to which the contraction principle will be applied. The underlying Banach space $E$ is the same as in the proof of Theorem 1, and the nonlinear operator $S$ is given by the right hand side of (18). In contrast to the linear case the operator $S$ is not defined on all of $E$. In the first part of the proof we shall describe a proper closed subset $F$ of $E$ which is mapped into itself by $S$. We define $F$ as the set of all functions $u \in E$ which satisfy the inequalities

$$|u(t,z)| \leqslant \rho \text{ and } |u_{z_k}(t,z)| \leqslant \frac{1}{\sqrt{d(t,z)}},$$

where $\rho = 2\eta\sqrt{d_0}(\beta + \gamma) < R$. Let $u \in F$ and $v = Su$. Then

$$|v_t(t,z)| \leqslant \sum_j |B_j||u_{z_j}| + |c| \leqslant \frac{\beta + \gamma}{\sqrt{d(t,z)}}.$$

Using the estimate

$$\int_0^{|t|} \frac{ds}{\sqrt{d(s,z)}} = -2\eta \sqrt{d(z) - \frac{s}{\eta}}\,\Bigg|_0^{|t|} \leqslant 2\eta\sqrt{d_0},$$

integration of $v_t$ with respect to $t$ yields

$$|v(t,z)| \leqslant (\beta + \gamma)2\eta\sqrt{d_0} = \rho.$$

Now we estimate the derivatives of $v$, using the notation $c(u) = c(t,z,u(t,z))$ and similarly for $B_j$. Nagumo's lemma implies

$$\left| \frac{\partial}{\partial z_k} c(u) \right| \leqslant \frac{\gamma C}{d^{3/2}}, \qquad |u_{z_j z_k}| \leqslant \frac{C}{d^{3/2}}$$

and

$$\left| \frac{\partial}{\partial z_k} B_j(u) \right| \leqslant \frac{\beta_j}{d},$$

where $d = d(t,z)$ and $C = C_{1/2} = 3\sqrt{3}/2$. Using the product rule of differentiation, we obtain

$$|v_{t,z_k}| \leqslant \frac{1}{d^{3/2}}(\beta C + \beta + \gamma C)$$

and, after integration with respect to $t$ (see (10))

$$|v_{z_k}| \leqslant \frac{2\eta}{\sqrt{d}}(\beta C + \beta + \gamma C) \leqslant \frac{1}{\sqrt{d}}.$$

The estimations above show that $u \in F$ implies $v = Su \in F$.

In the second part of the proof the difference $Su - Sv$ will be estimated for $u, v \in F$. In an abbreviated notation, we get

$$(Su - Sv)_t = \sum_j (B_j(u) - B_j(v)) u_{z_j} + \sum_j B_j(v)(u_{z_j} - v_{z_j}) + c(u) - c(v)$$

and

$$|(Su - Sv)_t| \leqslant \sum_j \beta_j' \frac{|u-v|}{d} + \sum_j \beta_j |u_{z_j} - v_{z_j}| + \frac{\gamma'}{d}|u-v|.$$

Since $|u - v| \leqslant \|u - v\|/d^p$ and $|u_{z_j} - v_{z_j}| \leqslant C_p\|u - v\|/d^{p+1}$, the estimates

$$|(Su - Sv)_t| \leqslant \frac{\|u-v\|}{d^{p+1}}\left(\beta' + \beta C_p + \gamma'\right)$$

and, after integration,

$$|Su - Sv| \leqslant \frac{\|u-v\|}{d^p}\frac{\eta}{p}\left(\beta' + C_p\beta + \gamma'\right)$$

are obtained. We multiply by $d^p$, take the supremum with respect to $G_\eta$, and arrive at the final estimate

$$\|Su - Sv\| \leqslant \frac{\eta}{p}\left(\beta' + C_p\beta + \gamma'\right)\|u - v\|.$$

This inequality shows that (for large $p$) $S$ is a contraction and hence concludes the proof. □

REMARK 1. Remark 4 of Section 2 carries over to the quasilinear case. The possibility that there are other solutions to the Cauchy problem not belonging to $E$ is carried ad absurdum by considering subregions of $G_\eta$.

REMARK 2. In a similar way, Remark 5 carries over. If $B_j$ and $c$ are real-valued for real values of $t$, $z$ and $u$, then the solution $u$ is real-valued for real $t$ and $z$. This is shown by successive approximation $u_{k+1} = Su_k$, starting with $u_0 = 0$.

REMARK 3. As in the linear case, the solution depends continuously on the initial value and on the right hand side of the differential equation. The proof is based on the estimate (15) and follows the same lines as before.

REMARK 4. The subset $F$ of $E$ constructed in the first part of the proof is easily seen to be compact. Hence the existence of a solution follows already from the first part of the proof, if Schauder's fixed point theorem is called upon. This proof would have some similarity with those in [**11**] and [**8**].

Let us note finally that the method used in this article is not confined to the classical C-K theorem. It can easily be adapted to abstract versions and, more generally, to functional differential equations of the Cauchy-Kowalevsky type, for example to delay-differential equations with delay in the $t$-direction (real case). In the case where the set $\Omega$ is the whole space $\mathbb{C}^n$, one may consider the space of entire holomorphic functions with finite norm

$$\|f\| = \sup|f(z)|e^{-\alpha|z|} \qquad (z \in \mathbb{C}^n, \alpha > 0).$$

In this space partial differentiation is a bounded operation, $\|\partial f/\partial z_i\| \leqslant \alpha e\|f\|$. Using this fact, we find that the Cauchy problem can be treated in a very simple way; see Redheffer and Walter [**19**] for more details.

## References

**1.** C. Bessaga, On the converse of the Banach "fixed-point principle", Colloq. Math., 7 (1959) 41–43.

**2.** A. Cauchy, Mémoire sur l'emploi du calcul des limites dans l'intégration des équations aux dérivées partielles. -Mémoire sur l'application du calcul des limites à l'intégration d'équations aux dérivées partielles. -Mémoire sur les systèmes d'équations aux dérivées partielles d'ordres quelconques, et sur leur réduction à des systèmes d'équations linéaires du premier ordre. Comptes Rend., 15 (1842) (Oeuvres complètes, 1. Série, Tome VII, 17-33, 33-49 and 52-58).

**3.** R. Courant and D. Hilbert, Methods of Mathematical Physics, vol. II, Interscience Publishers, 1962.

4. I. M. Gelfand and G. E. Silov, Generalized Functions, vol. 3: Theory of Differential Equations, Academic Press, New York, 1967.

5. J. Hadamard, Lectures on Cauchy's Problem in Linear Partial Differential Equations, Dover Publications, New York, 1952.

6. L. Hörmander, Linear Partial Differential Operators, Springer Verlag, Berlin-Heidelberg-New York, 1963.

7. F. John, Partial Differential Equations, 3rd ed., Springer Verlag, New York-Heidelberg-Berlin, 1978.

8. K. Keller and A. Schneider, Ein funktionalanalytischer Beweis des Satzes von Cauchy-Kowalewsky, Manuscripta math., 39 (1982) 31–37.

9. S. v. Kowalevsky, Zur Theorie der partiellen Differentialgleichungen, J. Reine Angew. Math., 80 (1) (1875) 1–32.

10. S. Mizohata, The Theory of Partial Differential Equations, Cambridge Univ. Press, 1973.

11. M. Nagumo, Über das Anfangswertproblem Partieller Differentialgleichungen, Japan. J. Math., 18 (1941-43) 41–47.

12. L. Nirenberg, An abstract form of the nonlinear Cauchy-Kowalewski theorem, J. Differential Geom., 6 (1972) 561–576.

13. T. Nishida, A note on a theorem of Nirenberg, J. Differential Geom., 12 (1977) 629–633.

14. L. V. Ovsjannikov, Singular operators in Banach scales, Dokl. Akad. Nauk SSSR, 163 (1965) 819–822; Soviet Math. Dokl., 6 (1965) 1025–1028.

15. ______, A nonlinear Cauchy problem in a scale of Banach spaces, Dokl. Akad. Nauk SSSR, 200 (1971) 789–792; Soviet Math. Dokl., 12 (1971) 1497–1502.

16. ______, Abstract form of the Cauchy-Kowalewski theorem and its applications (Russian). Partial Differential Equations (Proc. Conf. Novosibirsk, 1978), 88–94, 250 "Nauka" Sibirsk, Otdel., Novosibirsk, 1980.

17. T. H. Pate, A Direct Iterative Method for an Abstract Cauchy-Kowalewsky Theorem, Indiana Univ. Math. J., 30 (3) (1981) 415–425.

18. I. G. Petrovski, Lectures on Partial Differential Equations, Interscience Publishers 1961 (third printing).

19. R. Redheffer and W. Walter, Existence theorems for strongly coupled systems of partial differential equations over Bernstein classes, Bull. Amer. Math. Soc., 82 (1976) 899–902.

20. F. Treves, On the theory of linear partial differential operators with analytic coefficients, Trans. Amer. Math. Soc., 137 (1969) 1–20.

21. ______, An abstract nonlinear Cauchy-Kovalevska theorem, Trans. Amer. Math. Soc., 150 (1970) 77–92.

22. ______, Basic Linear Partial Differential Equations, Academic Press, New York-San Francisco-London, 1975.

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## THE GAMMA FUNCTION AND THE HURWITZ ZETA-FUNCTION

BRUCE C. BERNDT*

*Department of Mathematics, University of Illinois, Urbana, IL 61801*

The gamma function $\Gamma(x)$ may be defined by

$$\Gamma(x) = \lim_{n\to\infty} \frac{n!(n+1)^x}{x(x+1)\cdots(x+n)}, \tag{1}$$

where $x$ is any complex number, while the Hurwitz zeta-function $\zeta(s,x)$ is defined by

$$\zeta(s,x) = \sum_{k=0}^{\infty} (k+x)^{-s}, \tag{2}$$

*Research partially supported by National Science Foundation Grant No. MCS 7903359.

where $\sigma = \operatorname{Re} s > 1$ and $x$ is any complex number. (Normally, it is assumed that $0 < x \le 1$ in the definition of $\zeta(s,x)$.) Observe that $\zeta(s,1) = \zeta(s)$, where $\zeta(s)$ denotes the Riemann zeta-function. The function $\zeta(s,x)$ can be analytically continued into the entire complex $s$-plane and is holomorphic except for a simple pole at $s = 1$.

In 1894, Lerch [**9**, p. 13] established the following beautiful formula relating $\Gamma(x)$ and $\zeta(s,x)$.

THEOREM. *If the prime $'$ denotes differentiation with respect to $s$, then*

$$\operatorname{Log}\Gamma(x) = \zeta'(0,x) - \zeta'(0). \tag{3}$$

This formula is not particularly well known. However, it has achieved some prominence in recent research. If $L(s,\chi)$ denotes the Dirichlet $L$-function associated with the character $\chi$, then there is a classical formula for $L'(0,\chi)$ that depends upon (3). Analogues of this formula have been established for other $L$-functions, e.g., $p$-adic $L$-functions, and analogues of (3) naturally arise. (See, e.g., [**2**], [**3**], [**10**], and [**11**].)

The main purpose of this note is to present a new, short proof of (3) that is more elementary than previously known proofs [**9**], [**13**, pp. 59–60], [**14**, p. 271]. A second purpose is to show how (3) along with other properties of the Hurwitz zeta-function can be used to give short proofs of Gauss's multiplication theorem, Kummer's formula for $\operatorname{Log}\Gamma(x)$, and the reflection theorem. These proofs are new, but we do not claim that they are better or shorter than other proofs; we primarily wish to emphasize the little noticed connection between the two classical functions in the title of this paper.

*Proof of the Theorem.* For $\sigma > 1$ and $x > 0$, a straightforward application of the Euler-Maclaurin summation formula [**12**, p. 13] yields

$$\zeta(s,x) = \frac{x^{1-s}}{s-1} + \frac{x^{-s}}{2} - s\int_0^\infty \frac{t-[t]-\frac{1}{2}}{(t+x)^{s+1}}\,dt. \tag{4}$$

By analytic continuation, (4), in fact, is valid for $\sigma > -1$. Differentiating both sides of (4) and then setting $s = 0$, we find that

$$\begin{aligned} \zeta'(0,x) &= x\operatorname{Log} x - x - \frac{1}{2}\operatorname{Log} x - \int_0^\infty \frac{t-[t]-\frac{1}{2}}{t+x}\,dt \\ &= x\operatorname{Log} x - x - \frac{1}{2}\operatorname{Log} x + \lim_{n\to\infty}\left(-\int_0^n \frac{t+x}{t+x}\,dt\right. \\ &\quad \left.+\sum_{k=1}^{n-1} k\int_k^{k+1}\frac{dt}{t+x} + \left(\frac{1}{2}+x\right)\int_0^n \frac{dt}{t+x}\right) \\ &= \lim_{n\to\infty}\left(-x - n - \sum_{k=0}^{n}\operatorname{Log}(k+x) + \left(n+x+\frac{1}{2}\right)\operatorname{Log}(n+x)\right). \end{aligned} \tag{5}$$

In particular,

$$\zeta'(0) = \zeta'(0,1) = \lim_{n\to\infty}\left(-1 - n - \sum_{k=0}^{n}\operatorname{Log}(k+1) + \left(n+\frac{3}{2}\right)\operatorname{Log}(n+1)\right). \tag{6}$$

Thus, (5) and (6) give

$$\begin{aligned} \zeta'(0,x) - \zeta'(0) = \lim_{n\to\infty}&\left(1 - x + \sum_{k=0}^{n}\{\operatorname{Log}(k+1) - \operatorname{Log}(k+x)\}\right. \\ &\left.+\left(n+x+\frac{1}{2}\right)\operatorname{Log}(n+x) - \left(n+\frac{3}{2}\right)\operatorname{Log}(n+1)\right) \end{aligned} \tag{7}$$

$$= \lim_{n\to\infty}\left(1 - x + \sum_{k=0}^{n}\{\operatorname{Log}(k+1) - \operatorname{Log}(k+x)\}\right.$$

$$\left.+\left(n + x + \frac{3}{2}\right)\operatorname{Log}\left(\frac{n+x}{n+1}\right) - \operatorname{Log}(n+x) + x\operatorname{Log}(n+1)\right)$$

$$= \lim_{n\to\infty}\left(\sum_{k=0}^{n}\{\operatorname{Log}(k+1) - \operatorname{Log}(k+x)\}\right.$$

$$\left.-\operatorname{Log}(n+x) + x\operatorname{Log}(n+1)\right).$$

On the other hand, by (1),

$$\operatorname{Log}\Gamma(x) = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\operatorname{Log} k + x\operatorname{Log}(n+1) - \sum_{k=0}^{n}\operatorname{Log}(k+x)\right). \tag{8}$$

A comparison of (7) and (8) completes the proof for $x > 0$. By analytic continuation, (3) holds for all complex $x$.

In the sequel, we shall need one fact about $\Gamma(x)$ and a few properties of $\zeta(s,x)$. Recall that Euler's constant $\gamma$ may be defined by

$$\gamma = \lim_{n\to\infty}\left\{\sum_{k=1}^{n}\frac{1}{k} - \log n\right\}.$$

Then, from (1), it is an easy exercise to show that

$$\Gamma'(1) = -\gamma. \tag{9}$$

Recall that [**14**, p. 271]

$$\zeta(0,x) = \frac{1}{2} - x \tag{10}$$

and

$$\zeta'(0) = -\frac{1}{2}\operatorname{Log}(2\pi). \tag{11}$$

Lastly, we shall need the most important property of $\zeta(s,x)$, Hurwitz's formula [**12**, p. 37], [**14**, p. 269],

$$\zeta(s,x) = 2\Gamma(1-s)\left\{\sin(\pi s/2)\sum_{k=1}^{\infty}\frac{\cos(2\pi kx)}{(2\pi k)^{1-s}} + \cos(\pi s/2)\sum_{k=1}^{\infty}\frac{\sin(2\pi kx)}{(2\pi k)^{1-s}}\right\}, \tag{12}$$

where $0 < x < 1$ and $\sigma < 1$. (A simple proof of (12) can also be found in [**1**].)

GAUSS'S MULTIPLICATION THEOREM ([**14**, p. 240]). *For any complex number $x$ and positive integer $n$,*

$$\prod_{k=0}^{n-1}\Gamma\left(\frac{x+k}{n}\right) = (2\pi)^{\frac{1}{2}(n-1)}n^{\frac{1}{2}-x}\Gamma(x). \tag{13}$$

*Proof.* For $x > 0$ and $\sigma > 1$, by (2),

$$\zeta'(s,x) - \sum_{k=0}^{n-1}\zeta'\left(s,\frac{x+k}{n}\right) = \zeta'(s,x) + \sum_{k=0}^{n-1}\sum_{j=0}^{\infty}\frac{\operatorname{Log}\left(j+\frac{x+k}{n}\right)}{\left(j+\frac{x+k}{n}\right)^{s}} \tag{14}$$

$$= \zeta'(s,x) + \sum_{r=0}^{\infty} \frac{\operatorname{Log}\{(r+x)/n\}}{\{(r+x)/n\}^s}$$
$$= (1-n^s)\zeta'(s,x) - n^s \operatorname{Log} n \zeta(s,x),$$

where in the penultimate line we set $r = jn + k$. By analytic continuation, the extremal sides of (14) are equal for all $s$. Setting $s = 0$, employing (3), and using (10) and (11), we find that, for $x > 0$,

$$\operatorname{Log}\Gamma(x) - \sum_{k=0}^{n-1} \operatorname{Log}\Gamma\left(\frac{x+k}{n}\right) = \left(x - \frac{1}{2}\right)\operatorname{Log} n - \frac{1}{2}(n-1)\operatorname{Log}(2\pi),$$

which is clearly equivalent to (13). By analytic continuation, (13) holds for all complex $x$.

The following elegant representation is due to Kummer [7], [**8**, pp. 325–328]. Hardy [**4**], [**5**], [**6**, pp. 348–350, 428–432] has given two proofs, and references to other proofs may be found in [**4**]. See also [**14**, p. 250].

KUMMER'S FORMULA. *If* $0 < x < 1$, *then*

$$\operatorname{Log}\Gamma(x) = \frac{1}{2}\operatorname{Log}(\pi \csc(\pi x)) + \sum_{k=1}^{\infty} \frac{(\gamma + \operatorname{Log}(2\pi k))\sin(2\pi k x)}{\pi k}, \tag{15}$$

*where* $\gamma$ *denotes Euler's constant.*

*Proof.* Differentiate both sides of Hurwitz's formula (12) with respect to $s$, set $s = 0$, and employ (3) to find that

$$\begin{aligned} \operatorname{Log}\Gamma(x) &= \zeta'(0,x) - \zeta'(0) \\ &= -\Gamma'(1)\sum_{k=1}^{\infty}\frac{\sin(2\pi k x)}{\pi k} + \sum_{k=1}^{\infty}\frac{\cos(2\pi k x)}{2k} \\ &\quad + \sum_{k=1}^{\infty}\frac{\operatorname{Log}(2\pi k)\sin(2\pi k x)}{\pi k} - \zeta'(0). \end{aligned} \tag{16}$$

The second series on the right side of (16) is well known to equal $\frac{1}{2}\operatorname{Log}(\frac{1}{2}\csc(\pi x))$, $0 < x < 1$ [**14**, p. 190]. Using also (9) and (11), we readily find that (16) reduces to (15).

REFLECTION FORMULA ([**14**, p. 239]). *If $x$ is any complex number, then*

$$\Gamma(x)\Gamma(1-x) = \pi\csc(\pi x). \tag{17}$$

*Proof.* Suppose first that $0 < x < 1$. Then by Kummer's formula (15),

$$\operatorname{Log}\Gamma(x) + \operatorname{Log}\Gamma(1-x) = \operatorname{Log}(\pi\csc(\pi x)),$$

which is equivalent to (17). By analytic continuation, (17) holds for all $x$.

## References


**1.** B. C. Berndt, On the Hurwitz zeta-function, Rocky Mountain J. Math., 2 (1972) 151–157.

**2.** B. Ferrero and R. Greenberg, On the behavior of $p$-adic $L$-functions at $s = 0$, Invent. Math., 50 (1978) 91–102.

**3.** B. H. Gross and N. Koblitz, Gauss sums and the $p$-adic $\Gamma$-function, Ann. of Math., 109 (1979) 569–581.

**4.** G. H. Hardy, A new proof of Kummer's series for $\log\Gamma(a)$, Mess. Math., 31 (1902) 31–33.

**5.** ______, On Kummer's series for $\log\Gamma(a)$, Quart. J. Math., 37 (1906) 49–53.

**6.** ______, Collected papers, vol. IV, Clarendon Press, Oxford, 1969.

**7.** E. E. Kummer, Beitrag zur Theorie der Function $\Gamma(x) = \int_0^\infty e^{-v}v^{x-1}dv$, J. Reine Angew. Math., 35 (1847) 1–4.

**8.** ______, Collected papers, vol. II, Springer-Verlag, Berlin, 1975.

9. M. Lerch, Dalši studie v oboru Malmsténovských řad, Rozpravy Česke Akad. 3, no. 28, 1894, 63 pp.

10. C. J. Moreno, The Chowla-Selberg formula, J. Number Theory, 17 (1983) 226–245.

11. T. Shintani, On special values of zeta-functions of totally real quadratic fields, Proc. Inter. Congr. Math., Helsinki, 1978, vol. 2, Acad. Sci. Fennica, Helsinki, 1980, pp. 591–597.

12. E. C. Titchmarsh, The Theory of the Riemann Zeta-Function, Clarendon Press, Oxford, 1951.

13. A. Weil, Elliptic Functions According to Eisenstein and Kronecker, Springer-Verlag, Berlin, 1976.

14. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., University Press, Cambridge, 1962.

---

## 138. MISCELLANEA

Just as surely as our understanding of Nature is really valid only to the extent that it is mathematical, so also our understanding of higher domains must be based on mathematical models.

—R. Steiner (1861–1925), quoted in *Mathematical Reviews* 82k : 51001.

Each natural science contains genuine Science only to the extent that it contains Mathematics.... It may be that a formal philosophy of Nature in general, that is, a philosophy that deals only with general concepts, is possible without Mathematics, but a formal natural science dealing with definite objects (whether Physics or Psychology) is possible only with the use of Mathematics, and since each natural science contains only as much genuine Science as it contains *a priori* knowledge, it follows that a natural science is a genuine Science only to the extent that Mathematics can be applied to it.

—Immanuel Kant, Metaphysische Anfangsgründe der Naturwissenschaft, 1876.

*Perhaps we should sometimes listen to the other side. Goethe, for example, did not like Mathematics.*

Mathematicians are amazing people. In virtue of their accomplishments, they have set themselves up as a universal guild and will acknowledge only what suits their circle, what their method of organization can produce. A prominent mathematician once said, when someone strongly recommended a topic in Physics, "But can't it be reduced to calculation?"

—Goethe, Maximen und Reflexionen, no. 1277

It surely does not follow that the hunter who kills the game must also cook it. A cook might go hunting and shoot well; but he would be badly mistaken if he claimed that only a cook can be a good shot. It seems to me that this is the situation of mathematicians who claim that nobody can understand or discover physical phenomena without being a mathematician, since they should be pleased enough if the meat is brought to their kitchen for them to lard it with formulas and dress it as they like.

—Goethe, Maximen und Reflexionen, no. 1280

---

### ANSWER TO PHOTO ON PAGE 93

The photo is of Louis de Branges, who proved in 1984 the conjecture Bieberbach made in 1916, by establishing a strong inequality proposed by the Russian function-theorist I. M. Milin.

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S(13-16), L.** Problemes avec et sans...Problemes! Florentin Smarandache. Somipress, 1983, 175 pp, (P). Contains 140 original problems at the level of the Olympiad contests. Subject areas include arithmetic, trigonometry, geometry, analysis, algebra, and general logical analysis. LCL

**General, T(16-17), S, P, L.** Thinking, Problem Solving, Cognition. Richard E. Mayer. WH Freeman, 1983, xi + 426 pp, $16.95 (P). [ISBN: 0-7167-1441-8] A reorganized and updated version of Thinking and Problem Solving (TR, April 1978). This introduction to the psychology of cognition includes three additional chapters: "Intelligence," "Creativity Training," and "Mathematical Problem Solving." LCL

**General, S(13), L.** Unusual Mathematical Puzzles, Tricks, and Oddities. Matthew Mandl. Prentice-Hall, 1984, vi + 122 pp, $15.95. [ISBN: 0-13-938150-3] Mathematical and computer entertainments. Much familiar material mixed with some original. Three of the eight chapters are computer-related. On the light side. Uneven in content and in treatment. No special prerequisite for readers. Price is a bit steep. JK

**General, S*(13-15), L*.** The Lighter Side of Gravity. Jayant V. Narlikar. WH Freeman, 1982, viii + 194 pp, $10.95 (P). [ISBN: 0-7167-1344-6] A wonderfully clear popular exposition of gravitational theory from Aristotle to Hawking, enlivened with anecdotes and unobtrusive wry humor. Reminiscent of Gamow's classic 1,2,3,...Infinity, Narlikar's discourse--slightly mistitled--takes the lay reader on a revealing tour of contemporary cosmology, explaining in clear, simple language how black holes can radiate and binary starts can produce gravity waves. LAS

**Elementary, T*(13: 1), S.** Using Algebra. Ethan D. Bolker. Little, Brown & Co, 1983, xi + 298 pp, $22.95. [ISBN: 0-316-10114-1] Applied intermediate algebra, written to teach students to think about problems. Virtually all the examples and exercises are word problems (!): algebra is used to model and draw inferences about everyday real-world phenomena. The pace is leisurely and the problems are carefully stated and developed. The book teaches the use of the calculator in parallel with the mathematics. LCL

**Elementary, T(13: 1), S.** Elementary Algebra. C. Lee Welch, Gilbert M. Peter. Scott, Foresman, 1985, 390 pp, $23.95. [ISBN: 0-673-16576-0] Intended as a first course in algebra at the college level--contains a broad range of topics including a brief introduction to functions and complex numbers. Has a very straightforward approach and a high reading level. 4-6 well-chosen examples per section. Contains many applications and challenge problems--but thrust of text is the acquisition of skills. TR

**Education, T(14-15: 1, 2).** A Problem Solving Approach to Mathematics for Elementary School Teachers, Second Edition. Rick Billstein, Shlomo Libeskind, Johnny W. Lott. Benjamin/Cummings, 1984, xvi + 723 pp, $26.95. [ISBN: 0-8053-0856-3] Heuristic approach to problem solving. Changes from the First Edition (TR, November 1981; Extended Review, January 1982) include: inclusion of calculator and computer exercises; both LOGO and BASIC covered in chapter on computers. JRG

**History, P, L*.** Writings of Charles S. Peirce: A Chronological Edition, Volume 2, 1867-1871. Charles S. Peirce. Indiana U Pr, 1982, xlviii + 649 pp, $35. [ISBN: 0-253-37202-X] Continuation of a monumental work begun in Volume 1 (TR, February 1983); a landmark edition including numerous manuscripts never before published. Contains extensive scholarly and editorial notes, three introductions, and a complete list of Peirce manuscripts. LAS

**Foundations, S(18), P.** Mathematical Applications of Category Theory. Ed: J.W. Gray. Contemporary Math., V. 30. AMS, 1984, vii + 307 pp, $28 (P). [ISBN: 0-8218-5032-6] Introduction and ten papers dealing with some aspect of category theory ranging from logic to differential geometry to topos theory. Separate sets of references. JC

Combinatorics, T(17: 1), S, P, L. The Theory of Uniform Distribution. Edmund Hlawka. Transl: Henry Orde. AB Academic, 1984, x + 141 pp. [ISBN: 0-907360-02-5] An appealing introduction to the subject in three parts. The first gives the general foundations, the second introduces the abstract theory, and the third applications to numerical analysis, analysis and number theory. Originally published in German in 1979. Good list of references. No exercises. CEC

Number Theory, T*(14-15: 1, 2), S, P, L*. Topics from the Theory of Numbers, Second Edition. Emil Grosswald. Birkhauser Boston, 1984, xv + 333 pp, $29.95. [ISBN: 3-7643-3044-9] The first half of this book is an outstanding introduction to elementary number theory which is essentially the same as the First Edition (TR, May 1976; Extended Review, May 1968). The second half consists of topics from analytic and algebraic number theory and has been thoroughly redone for this new edition. Interesting selection of topics and good exercises. CEC

Number Theory, S(17), P*, L. Lecture Notes on Primality Testing and Factoring: A Short Course at Kent State University. Carl Pomerance. MAA, 1984, 34 pp, $3.50 (P). [ISBN: 0-88385-054-0] A brief monograph which is a superb introduction to state-of-the-art factoring and primality testing algorithms. Includes a list of references. CEC

Linear Algebra, T(16-17: 2), S. Lineare Algebra und Geometrie. W. Klingenberg. Hochschultext. Springer-Verlag, 1984, xi + 313 pp, $12.60 (P). [ISBN: 0-387-13427-1] A fairly elementary but comprehensive introduction to linear algebra over an arbitrary commutative field, together with a development of the elements of affine, Euclidean, projective and non-Euclidean geometry based on linear algebra. No exercises. JD-B

Linear Algebra, T(13-15: 1), S, L. Linear Algebra, Second Edition. Larry Smith. Undergrad. Texts in Math. Springer-Verlag, 1984, vii + 362 pp, $29.95. [ISBN: 0-387-96015-5] The most significant change from the First Edition (TR, December 1978) is the inclusion of a new chapter on the Jordan canonical form. JS

Linear Algebra, T(14-17: 1), L. Linear Algebra: An Introductory Approach. Charles W. Curtis. Undergrad. Texts in Math. Springer-Verlag, 1984, x + 337 pp, $24. [ISBN: 0-387-90992-3] A Fourth Edition with no major changes from the Third Edition published by Allyn and Bacon. (1968 Second Edition, TR, October 1969; Extended Review, November 1972; 1974 Third Edition, TR, October 1974.) JS

Group Theory, T(17-18: 1, 2), S. Lie Groups, Lie Algebras, and Their Representations. V.S. Varadarajan. Grad. Texts in Math., No. 102. Springer-Verlag, 1984, xiii + 430 pp, $29.80. [ISBN: 0-387-90969-9] Relatively self-contained treatment of the algebraic and analytic aspects of Lie group theory, presuming only some knowledge of differentiable manifolds and topological groups. Culminates in detailed development of the structure and representation theory of semisimple Lie groups and Lie algebras with both Harish-Chandra's algebraic methods and Weyl's transcendental methods via compact groups. Many nice exercises develop aspects of the theory not covered in the text. RM

Group Theory, S, P*, L. The Classification of Finite Simple Groups: Volume 1: Groups of Noncharacteristic 2 Type. Daniel Gorenstein. Univ. Ser. in Math. Plenum Pr, 1983, x + 487 pp, $59.50. [ISBN: 0-306-41305-1] A comprehensive outline of the classification of simple groups of noncharacteristic 2 type. Gorenstein is a guide and companion along the way, providing direction, overview, intuition, and purpose. The mathematical community is indeed fortunate to receive this invaluable contribution to the literature. LCL

Group Theory, T(16-18: 1, 2). An Introduction to the Theory of Groups, Third Edition. Joseph J. Rotman. Allyn & Bacon, 1984, x + 422 pp, $40.71. [ISBN: 0-205-07963-6] An updated and reorganized edition that represents some significant new material and changes in older material. New material is found on permutations, Mathieu groups, automorphisms of S6, the fundamental theorem of projective geometry and much more. The treatment of amalgams and HNN extensions is now based on pushouts and the Seifert-van Kampen theorem. (Second Edition, TR, March 1974.) JAS

Group Theory, P. FC-groups. M.J. Tomkinson. Research Notes in Math., No. 96. Pitman, 1984, 171 pp, $19.95 (P). [ISBN: 0-273-08566-2] A group G is an FC-group if every conjugacy class is finite. This volume gives an account of results of Hall and Gorčakov about embedability of certain classes of FC-groups into products of finite groups, and surveys advanced techniques. Twenty-five unsolved questions; sizable bibliography. RB

Algebra, P. Classgroups of Group Rings. Martin Taylor. London Math. Soc. Lect. Note Ser., No. 91. Cambridge U Pr, 1984, xiii + 119 pp, $17.95 (P). [ISBN: 0-521-27870-8] A number of new approaches and techniques for investigating modules which are locally free over an integral group ring, and the consequent problem of determining whether or not the module is globally free. LCL

Algebra, P. Characters of Reductive Groups Over a Finite Field. George Lusztig. Princeton U Pr, 1984, xxi + 384 pp, $50; $19.50 (P). [ISBN: 0-691-08350-9] The main result of this work is an explicit formula for the multiplicities with which the various irreducible representations of G(Fq) appear in the virtual representation of G(Fq) where G is a reductive algebraic group with connected center defined over a finite field, Fq. Extensive bibliography. CEC

Algebra, S(17), P. The Umbral Calculus. Steven Roman. Pure & Appl. Math., V. 111. Academic Pr, 1984, x + 193 pp, $35. [ISBN: 0-12-594380-6] Elementary introduction to modern umbral calculus with some applications and a short chapter on nonclassical umbral calculi. Clear, friendly writing style

provides excellent overview of subject as it is developed. No problems. BH

**Algebra, S(17), P.** Stone Spaces. Peter T. Johnstone. Stud. in Adv. Math., No. 3. Cambridge U Pr, 1982, xxi + 370 pp, $59.50. [ISBN: 0-521-23893-5] Presents Stone's representation theorem for Boolean algebras and some of its consequences in topology, ring theory, sheaf theory, and lattice theory. Includes an introduction with a description of the historical development of the theorem and its consequences. Historical notes at the end of each chapter and an extensive bibliography. BH

**Algebra, S(18), P.** Relative Finiteness in Module Theory. Toma Albu, Constantin Năstăsescu. Pure & Appl. Math.,V. 84. Dekker, 1984, xii + 190 pp, $39.75. [ISBN: 0-8247-7143-5]

**Algebra, P.** FPF Ring Theory: Faithful Modules and Generators of mod-R. Carl Faith, Stanley Page. London Math. Soc. Lect. Note Ser., No. 88. Cambridge U Pr, 1984, 168 pp, $19.95 (P). [ISBN: 0-521-27738-8] Based on lecture notes by Faith, it is a systematic study of the known results concerning finitely pseudo-Frobenius (FPF) rings. Essentially self-contained; basics lead to discussion of semi-perfect, semi-prime, non-singular, Goldie, and self-injective FPF rings. Open questions; bibliography; index. JS

**Calculus, T(13: 1).** A Short Calculus: An Applied Approach, Fourth Edition. Daniel Saltz. Scott, Foresman, 1985, 616 pp, $26.95. [ISBN: 0-673-16625-2] Contains a good general overview of algebra. Many straightforward examples and applications in business, economics, and life, physical, and social sciences. Very broad-based text without much background material. Stresses recognition of many types of notation within each topic. (First Edition, TR, June-July 1973; Second Edition, TR, October 1977; Third Edition, TR, February 1981.) TR

**Calculus, T*(14-16: 1, 2), S*, L*.** Multivariable Analysis. G. Baley Price. Springer-Verlag, 1984, xiv + 655 pp, $39. [ISBN: 0-387-90934-6] A new approach (triggered many years ago by a penetrating question by a bright student in a course in analytic geometry to the theory of differentiation for functions which map sets in Rn into Rm. Reduces to the standard theory in the one-dimensional case. Includes Riemann integration and related topics in analysis. Methods are based on oriented simplicial complexes and determinants. Numerous examples and exercises. Recommmended for a different kind of course in multivariable analysis. Schools with honors courses should have a look. For the student "with a first course in calculus and the ability to read and understand mathematical definitions, theorems and proofs." But beware: even the author cautions "it is not always easy." JK

**Calculus, S(13).** Problem Book for First Year Calculus. George W. Bluman. Problem Books in Math. Springer-Verlag, 1984, xv + 385 pp, $39. [ISBN: 0-387-90920-6] What should the speed limit be to maximize traffic flow during rush hour? 1000 problems from single variable calculus, emphasizing applications but covering as well standard techniques and theory. Each section opens with a brief synopsis of terms, followed first by solved problems and then by supplementary problems for which only answers are provided. Problems are drawn from university exams throughout Canada. LAS

**Real Analysis, S(18), P.** Intertwining Functions on Compact Lie Groups. B. Hoogenboom. CWI Tract No. 5. Math Centrum, 1984, 86 pp, Dfl. 13,20 (P). [ISBN: 90-6196-271-4] Based on work by Koornwinder, James, Constantine, and Vretare, this book explores some of the relationships between orthogonal polynomials and intertwining functions on a compact Lie group. References, index. JS

**Complex Analysis, S(17), P*.** The Cauchy Method of Residues: Theory and Applications. Dragoslav S. Mitrinović, Jovan D. Kečkić. Math. & Its Applic. D Reidel, 1984, xiv + 361 pp, $67. [ISBN: 90-277-1623-4] Devoted entirely to residues in one complex variable, but briefly mentions several complex variables. Contains a wealth of applications to ordinary differential equations, partial differential equations, finite differences, special functions, and an impressive collection of contour integrals. Many bibliographic items, but few are recent. Pleasant biography of Cauchy's life and work. No exercises, but great supplementary reading or reference. YN

**Complex Analysis, T(17: 1), P.** Funktionentheorie I. Reinhold Remmert. Grund. Math., No. 5. Springer-Verlag, 1984, xiii + 324 pp, $16.10 (P). [ISBN: 0-387-12782-8] A modern introduction to complex analysis, with many historical notes but almost no exercises. JD-B

**Differential Equations, T*(16-17), S, L.** Introduction to Partial Differential Equations with Applications. Mark A. Pinsky. McGraw-Hill, 1984, x + 326 pp, $32.95. [ISBN: 0-07-050117-3] For students who have had a first course in ordinary differential equations. Stress on systematic solution algorithms based on separation of variables and Fourier series and integrals. Includes classical techniques, plus chapters on asymptotic solutions and numerical solutions. Numerous worked examples, good exercises, physical motivation and applications. LCL

**Differential Equations, P.** Lectures on Stochastic Differential Equations and Malliavin Calculus. S. Watanabe. Springer-Verlag, 1984, 111 pp, $7.30 (P). [ISBN: 0-387-12897-2] Mathematical disquisition on stochastic differential equations arising from parabolic differential operators. Concise introduction followed by one hundred pages of definitions, lemmas, theorems, and computations. YN

**Partial Differential Equations, T(17: 1), P.** Lectures on Three-Dimensional Elasticity. P.G. Ciarlet. Springer-Verlag, 1983, 149 pp, $7.90 (P). In the first chapter a nonlinear system of partial differential equations is established as a mathematical model of elasticity. In the second (and last), questions of existence of solutions are studied. Tools are the implicit function theorem and the theory of J.M. Ball. Some exercises. References. Some open problems. JK

**Numerical Analysis**, T?(17), S. Numerical Solution of Differential Equations, Second Edition. M.K. Jain. Halsted Pr, 1984, xix + 698 pp, $34.95. [ISBN: 0-470-27389-5] Encyclopedia of numerical methods for approximating solutions to differential equations. Presumes some knowledge of advanced calculus and elementary numerical analysis. Added problems and enlarged chapter on the finite element method distinguish text from its First Edition (TR, October 1980). Large number of problems with solutions in back. Although text is rather dry, the bibliographic references after each chapter and extensive bibliography make it a useful reference. BH

**Functional Analysis**, T(18: 1, 2), S, P. Applied Nonlinear Analysis. Jean-Pierre Aubin, Ivar Ekeland. Wiley, 1984, xi + 518 pp, $47.50. [ISBN: 0-471-05998-6] Topics in nonlinear functional analysis--solving nonlinear equations and inclusion relations, variational problems, non-smooth analysis--are studied together with applications, e.g., to game theory, mathematical economics, convex optimization. Authors assume only the most elementary functional analysis; they prove many classical results, especially of convex analysis. Informal, readable--directed at variety of consumers of mathematics. PZ

**Functional Analysis**, P. The Analogue of the Group Algebra for Topological Semigroups. H.A.M. Dzinotyiweyi. Research Notes in Math., No. 98. Pitman, 1984, 196 pp, $19.95 (P). [ISBN: 0-273-08610-3] "A great deal of abstract harmonic analysis on a locally compact topological group is pivoted at the object L1(G)--the so-called group algebra of G. The purpose of this book is to present analogues of L1(G) for topological semigroups..." For researchers and students in the specialty, and possibly theoretical probabilists, functional analysts. RB

**Functional Analysis**, T(18: 1), S, P. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions. Christian Berg, Jens Peter Reus Christensen, Paul Ressel. Grad. Texts in Math., No. 100. Springer-Verlag, 1984, x + 289 pp, $39. [ISBN: 0-387-90925-7] Centered on the general theme of positive definite functions defined as Abelian semigroups with involution. Treatment is largely self-contained, proceeding through preliminary measure theory to such topics as Schoenberg's theorem, moment functions, and Hoeffding's inequality. Notes and remarks; bibliography; index. JS

**Functional Analysis**, P. Spaces of Measures. Corneliu Constantinescu. Stud. in Math., No. 4. Walter de Gruyter, 1984, 444 pp, DM 128. [ISBN: 3-11-008784-7] An attempt to present a systematic approach to the theory of spaces of measures which has grown out of initial work by Nikodym, Orlicz, Pettis, Vitali, Hahn, Saks, and Phillips. Applications to locally convex vector lattices. References, index. JS

**Analysis**, P. Operator Commutation Relations. Palle E.T. Jørgensen, Robert T. Moore. Math. & Its Applic. D Reidel, 1984, xviii + 493 pp, $69. [ISBN: 90-277-1710-9] The authors use certain operator commutation relations to unify and extend recent results in the theory of Lie algebras. In particular, they study families of noncommuting unbounded operators on Banach spaces by techniques that generalize the classical Lie theory of matrices and Lie transformation groups. LCL

**Analysis**, S(18), P*. Tata Lectures on Theta II: Jacobian Theta Functions and Differential Equations. David Mumford. Birkhauser Boston, 1984, xiv + 272 pp, $22. [ISBN: 0-8176-3110-0] The second in a series of three volumes which survey the theory of theta functions. Presents an explicit elementary construction of hyperelliptic Jacobian varieties and a self-contained introduction to the theory of Jacobians. CEC

**Algebraic Geometry**, P. Topics in Transcendental Algebraic Geometry. Ed: Phillip Griffiths. Annals of Math. Stud., No. 106. Princeton U Pr, 1984, viii + 316 pp, $45; $12.50 (P). [ISBN: 0-691-08335-5; 0-691-08339-8] The proceedings of a seminar (1981-82, Institute for Advanced Study) to explore ways formal Hodge theory might be applied in algebraic geometry. Seventeen lectures aim to give an intuitive picture of what is known and of some of the available techniques. LCL

**Differential Geometry**, P. Mathematical Essays in Honor of Su Buchin. Ed: C.C. Hsiung. Heyden & Son, 1983, viii + 276 pp, $35. [ISBN: 9971-950-98-7] 26 research papers in differential geometry, reprinted from the Journal of Differential Geometry and the Chinese Annals of Mathematics, on the occasion of the 80th birthday of Su Buchin, President of Fudan University in Shanghai. LAS

**Geometry**, S(13), P, L*. Orderly Tangles: Cloverleafs, Gordian Knots, and Regular Polylinks. Alan Holden. Columbia U Pr, 1983, x + 97 pp, $19.95. [ISBN: 0-231-05544-7] A descriptive work which includes photographs of remarkable models of various highway interchanges, knots, cat's cradles, chained polylinks and regular polylinks. Informative and aesthetically pleasing. CEC

**Geometry**, S*(13-16), L**. Geometry in Architecture. William Blackwell. Wiley, 1984, xii + 185 pp, $34.95. [ISBN: 0-471-09683-0] Reflects the author's "lifelong fascination with the endless relationships of geometric forms and the continual discovery of their applications to architecture and design." No theorems but lots of illustrations. Excellent drawings and magnificant (B & W) photographs with cogent comments. The reader is led from regular polygons in the plane to the classical solids in three dimensions. A book to be savored by all. No special knowledge of geometry is required. Read--or just browse--and enjoy this delightful book. JK

**Algebraic Topology**, T(18: 2), P. Lecture Notes in Mathematics-1053: Nilpotente Gruppen und nilpotente Räume. Peter J. Hilton. Springer-Verlag, 1984, v + 219 pp, $11 (P). [ISBN: 0-387-12910-3]

**Algebraic Topology, P.** Topological Topics. I.M. James. London Math. Soc. Lect. Note Ser., No. 86. Cambridge U Pr, 1983, 184 pp, $21.95 (P). [ISBN: 0-521-27581-4] "Articles on algebra and topology presented to P.J. Hilton in celebration of his sixtieth birthday." Includes a list of publications, two survey articles, and a number of papers dedicated to Hilton. JAS

**Algebraic Topology, P.** Lecture Notes in Mathematics-1069: Bordism of Diffeomorphisms and Related Topics. Matthias Kreck. Springer-Verlag, 1984, 144 pp, $7.50 (P). [ISBN: 0-387-13362-3] The bordism relation is defined on the class of all diffeomorphisms of all manifolds of fixed dimension m, and is based on the notion of bordism of m-manifolds. Bordism groups of diffeomorphisms, whose elements are bordism classes, are computed and applications are given in this research monograph. RB

**Statistics, S(17).** Statistical Methods and the Improvement of Data Quality. Ed: Tommy Wright. Academic Pr, 1983, xvii + 357 pp, $25. [ISBN: 0-12-765480-1] A set of papers presented at the Small Conference on the Improvement of the Quality of Data in November, 1982 at Oak Ridge, Tennessee. Most of the papers deal with large data collection systems and are concerned with survey sampling topics. MT

**Statistics, T*(15: 2).** Elementary Statistics for Business and Economics. Carl-Louis Sandblom. Walter de Gruyter, 1983, 352 pp, $21.80. [ISBN: 3-11-008302-7] This text assumes knowledge of differential and integral calculus and is intended for basic statistics courses in business administration and economics. It is a relatively brief text but its topics still include summary statistics, probability, random variables, sampling, estimation and hypothesis testing, analysis of variance, experimental design, regression, and time series. MT

**Statistics, P.** Percentage Points of the Distribution of the Correlation Coefficient. Dolun Öksoy, Leo A. Aroian. Inst. of Admin. & Manag. (Union Coll. & Univ., Schenectady, NY 12308), 1982, vi + 128 pp, (P). Tables of the percentage points of the distribution of r for sample sizes from three to 500 and population correlations from zero to one. MT

**Statistics, P.** Tables of the Distribution of the Correlation Coefficient. Dolun Öksoy, Leo A. Aroian. Inst. of Admin. & Manag. (Union Coll. & Univ., Schenectady, NY 12308), 1982, xxxiii + 132 pp, (P). Tables of the cumulative distribution function for r, the Pearson product moment coefficient of correlation, for sample sizes varying from three to 500 and for population correlations varying from zero to one. MT

**Statistics, P.** Probability, Statistics and Analysis. Ed: J.F.C. Kingman, G.E.H. Reuter. London Math. Soc. Lect. Note Ser., No. 79. Cambridge U Pr, 1983, viii + 286 pp, $29.95 (P). [ISBN: 0-521-28590-9] A collection of research papers dedicated to David Kendall on his 65th birthday and spanning his interests in mathematical statistics, probability, and analysis. LCL

**Statistics, S*(15-18), P*, L.** S: An Interactive Environment for Data Analysis and Graphics. Richard A. Becker, John M. Chambers. Stat./Prob. Ser. Wadsworth, 1984, xii + 550 pp, $23.95 (P). [ISBN: 0-534-03313-X] An appealing introduction to Bell Labs' UNIX-based statistical package S, an interactive system emphasizing graphical and exploratory methods. Simple one- or two-line commands yield dramatic visualizations of data, including scatterplots, starplots, boxplots, grid plots, cluster plots, and contour plots. Half of the book illustrates the use of S with numerous real data sets; the other half provides reference details on all S commands. LAS

**Statistics, P.** Statistics and Probability. Ed: J. Mogyoródi, I. Vincze, W. Wertz. D Reidel, 1984, x + 415 pp, $45.75. [ISBN: 90-277-1675-7] Proceedings of the third Pannonian Symposium on Mathematical Statistics held at Visegrád, Hungary, September 13-18, 1982. JAS

**Statistics, P*, L.** Statistical Computation. J.H. Maindonald. Wiley, 1984, xviii + 370 pp, $39.95. [ISBN: 0-471-86452-8] This book encourages the use of statistical software based on modern algorithms developed by numerical analysts. Many of these algorithms are discussed. About two-thirds of the book is devoted to multiple linear regression and closely related topics. AO

**Statistics, S(16), P.** Analyse des Données. Francis Cailliez. Pr U Montreal, 1984, 103 pp, $12 (P). [ISBN: 2-7606-0657-0] Author's lecture notes from 1982 NATO seminar; in French. Interprets concepts of multivariate data analysis (mean, correlation, factor...) in terms of linear algebra (projection, angle, eigenvector...). The main novelty lies in using the inner product of linear operators in this context. (However, similar results appeared first in M.L. Eaton, Multivariate Statistics, Wiley, 1983. TR, August-September 1984.) Concise, well written. YN

**Computer Literacy, S(13-15), L*.** The Computer Comes of Age: The People, the Hardware, and the Software. R. Moreau. Transl: J. Howlett. Ser. in History of Comp. MIT Pr, 1984, x + 227 pp, $19.95. [ISBN: 0-262-13194-3] A concise history of computing used as a vehicle for explaining the nature of modern computing. Focuses particularly on the development of mainframe machines and languages, up to the late 60's; does not get into the microcomputer age. LAS

**Computer Literacy, S(13-15), L.** Are Computers Alive? Evolution and New Life Forms. Geoff Simons. Birkhauser Boston, 1983, xi + 212 pp, $14.95 (P). [ISBN: 0-8176-3144-5] A fascinating, persuasive, well-researched examination of the birth of species machina sapiens. Beginning with issues in the definition of life (biochemistry vs. information theory), Simons examines behavioral, anatomical, and psychological features of computers that, in his view, support a hypothesis that computers are alive. LAS

**Computer Literacy, S(13).** The 3-D Animated Apple. Phil Cohen. Prentice-Hall, 1983, viii + 198 pp, $18.95 (P). [ISBN: 0-13-920224-2] Text is out of step with current technology. Intended as an inspiration--but those who might be interested will not be challenged by the information presented. Centered around two general programs--one 2-D graphics and one 3-D graphics program; but contains other related examples. TR

**Computer Literacy, P*.** The Art of Computer Conversation: A New Medium for Communication. Brian R. Gaines, Mildred L.G. Shaw. Prentice-Hall, 1984, x + 214 pp, $16.95 (P). [ISBN: 0-13-047332-4] Advice and examples (both good and bad) of human-machine dialogue, organized into 10 chapters and 30 proverbs (e.g., "Be consistent"). Much of the advice is routine, albeit too often ignored. LAS

**Computer Programming, P*, L.** The American Pascal Standard with Annotations. Henry Ledgard. Springer-Verlag, 1984, 97 pp, $16.95 (P). [ISBN: 0-387-91248-7] A two-column presentation of the ANSI standard for Pascal, with commentary and examples paralleling the formal statement of the standard. Employs several type fonts and well-conceived layout "to help reduce the mental effort needed to understand the standard." LAS

**Computer Programming, T.** Learning BASIC Programming Essentials. Carl Grame, Dan O'Donnell. SRA, 1984, xiv + 282 pp, $16.95 (P). [ISBN: 0-574-21370-8] A self-paced text which assumes the essentials of elementary algebra. Designed as system independent--should not be used with a micro-based system. Has many consumer applications. Provides a list of possible errors and corrections. TR

**Computer Programming, T*(13: 1).** Programming in Pascal, Second Edition. Peter Grogono. Addison-Wesley, 1984, xi + 420 pp, $21.95 (P). [ISBN: 0-201-12070-4] Primarily a language manual. Covers the entire language systematically. Numerous examples of complete programs are given. The basic structure and tone are the same as in the First Edition (TR, February 1979). AO

**Computer Programming, S(13-15), L.** Programming the IBM Personal Computer: Assembly Language. Chao C. Chien. Holt, Rinehart & Winston, 1984, xiv + 299 pp, $18.45 (P). [ISBN: 0-03-070442-1] For picking up IBM PC assembly on one's own. Presumes no programming experience and only rudiments of PC operation (e.g., copying a diskette). Use of text editor, internal computer organization, mention of top-down approach and structured programming. Library of useful low-level subroutines. A good choice for self study. RB

**Computer Programming, P.** Lecture Notes in Computer Science-167: International Symposium on Programming. Ed: M. Paul, B. Robinet. Springer-Verlag, 1984, vi + 262 pp, $13.50 (P). [ISBN: 0-387-12925-1] Twenty-two papers in French and English, the proceedings of the sixth biennial symposium held in Toulouse in April, 1984. Topics include concurrency and optimization, e.g., the vectorization language Vesta for high performance scientific computation; data network protocols, e.g., application of temporal logic to design of protocols; miscellaneous topics. RB

**Computer Programming, T(14-15: 1), S, P, L.** Using the IBM Personal Computer: Organization and Assembly Language Programming. Mark A. Franklin. Holt, Rinehart & Winston, 1984, xv + 357 pp, $20.45 (P). [ISBN: 0-03-062862-8] Comprehensive introduction to assembly on the IBM PC: nearly 30% of text devoted to computer organization, architecture, etc., before assembly language instructions. Presumes experience with algorithm formation, coding in high level languages. Advanced topics; bibliography; numerous exercises. Somewhat technical for relatively unsophisticated students; good choice for the experienced programmer. RB

**Software Systems, S?(17).** Program Transformation and Programming Environments. Ed: Peter Pepper. NATO ASI Ser. F, V. 8. Springer-Verlag, 1984, xiv + 378 pp, $39.50. [ISBN: 0-387-12932-4] Includes discussions and position papers from an Advanced Research Workshop of the NATO Science Committee. Deals with formal methods in software engineering and the application and acceptance of these methods. FA

**Software Systems, S, L.** MS-DOS User's Guide. Paul Hoffman, Tamara Nicoloff. Osborne/McGraw-Hill, 1984, viii + 312 pp, $17.95 (P). [ISBN: 0-88134-131-2] An introduction to versions 1 and 2 of Microsoft's Operating System MS-DOS, the generic version of IBM's PC-DOS. Explains how to use each feature, including the UNIX-like pipe, redirect, and directory features of Version 2, but does not provide any details of interest to those writing programs that interact with the operating system. LAS

**Computer Science, T(13-14: 1).** Structure and Interpretation of Computer Programs. Harold Abelson, Gerald Jay Sussman, Julie Sussman. Elect. Engin. & Comp. Sci. Ser. MIT Pr, 1985, xx + 542 pp, $30. [ISBN: 0-262-01077-1] This is an introductory computer science text which approaches its subject matter in a very different, and quite sophisticated, manner. It treats programming and languages as symbol manipulation systems and concentrates almost exclusively on problemm solving and high level abstraction mechanisms. The language that is used to illustrate examples is LISP--a functional language used frequently in artificial intelligence. This is an introductory text appropriate to only the most capable and mathematically sophisticated student. MS

**Computer Science, P.** Analysis and Simulation of Semiconductor Devices. Siegfried Selberherr. Springer-Verlag, 1984, xiv + 294 pp, $54. [ISBN: 0-387-81800-6] The book discusses techniques for constructing formal models of semiconductor devices--i.e., chips, circuit boards, and microprocessors. Using these models the designer can then study the behavior and stability of that device. The models discussed are mathematical simulation models. The author shows the development of the model and also gives techniques for solving the resulting equations. The mathematics is quite high

level and requires an understanding of calculus and linear algebra. The final chapter presents some results of the model when used to study one class of semiconductors called MOSFETs. MS

**Computer Science, S(17-18).** Lecture Notes in Computer Science-173: Semantics of Data Types. Ed: G. Kahn, D.B. MacQueen, G. Plotkin. Springer-Verlag, 1984, vi + 391 pp, $16.50 (P). [ISBN: 0-387-13346-1] Includes selected papers from the International Symposium on the Semantics of Data Types. Discusses a variety of data typing notations and mechanisms from the viewpoint of language designers as well as of pure and applied logicians. FA

**Computer Science, S(14-15).** On-Line Systems Design and Implementation (Using COBOL and Command Level CICS). Charles J. Kacmar. Reston, 1984, x + 404 pp, $26.95. [ISBN: 0-8359-5231-2] Most new applications of computers are on-line--that is, the user is connected directly to the computer. Because of this direct interaction between the computer and a (possibly inexperienced) user, on-line systems must be designed in a different way. Extreme care must be taken in the design of the user interface, the screen format, on-line assistance, error handling, and the use of graphics. This text addresses techniques for the most effective design of interactive, on-line systems. MS

**Computer Science, S(15-16).** Portability and Style in Ada. Ed: John Nissen, Peter Wallis. Cambridge U Pr, 1984, xii + 202 pp, $24.95. [ISBN: 0-521-26482-0] Ada is a new system implementation language which is a standard for all new computer systems developed by the Department of Defense. It is a very powerful language incorporating many new features not found in existing languages like Pascal, FORTRAN, or C. This text is a guide to using Ada in a correct and elegant way. It discusses how to make optimal use of such new features as packages, tasks, and data abstraction. MS

**Computer Science, S(15-16), P.** The 8087 Primer. John F. Palmer, Stephen P. Morse. Wiley, 1984, viii + 182 pp, $16.95 (P). [ISBN: 0-471-87569-4] Two parts on generalities of data representation and the ideas of coprocessors, and on 8087 architecture, precede a third section on programming. A hybrid book, part computer science and part processor manual. JAS

**Computer Science, P.** Programmable Assembly. Ed: Wilfred B. Heginbotham. Intern. Trends in Manufacturing Tech. Springer-Verlag, 1984, 349 pp, $43. [ISBN: 0-387-13479-4] Discusses the use of robots--i.e., mobile, flexible, and programmable devices--in production and control. It is a collection of about 30 papers which address a wide range of issues in the area of robotics, including robot design, applications, and economic and social concerns. The reader should have a working knowledge of both computer science and mechanical engineering to appreciate the relatively high-level discussions. MS

**Computer Science, L*.** Handbook of Algorithms and Data Structures. G.H. Gonnet. Internat. Comp. Sci. Ser. Addison-Wesley, 1984, xi + 286 pp, $25.95 (P). [ISBN: 0-201-14218-X] A reference work on algorithms and data structures for their implementation. Covers searching, sorting, selection, and arithmetic algorithms. Theoretical complexity results and implementations in both C and Pascal are given for each algorithm of practical importance. AO

**Computer Science, P.** Local Area Networks: Issues, Products, and Developments. V.E. Cheong, R.A. Hirschheim. Ser. in Comp. Wiley, 1983, xv + 190 pp, $29.95. [ISBN: 0-471-90134-2] A concise introduction to management issues, engineering concepts, and hardware options for LANs--networks that connect computers within distances of .1 to 10 km at transmission rates of .1 to 10 Mbps (million bits per second). Discusses various types (e.g., Ethernet, Cambridge Ring) as well as alternatives (e.g., private automatic branch exchange--PABX). LAS

**Computer Science, T(15-16), P.** Microcomputer System Design: An Advanced Course. Ed: M.J. Flynn, N.R. Harris, D.P. McCarthy. Springer-Verlag, 1984, vii + 397 pp, $19.95 (P). [ISBN: 0-387-13545-6] This text represents the course notes from an advanced course in microcomputer design held at Trinity College in July, 1981. The course addresses current technology in LSI/FLSI design, hardware structures, and language design. The course is a series of lectures by 8 professionals in the field of computer architecture and design. MS

**Computer Science, T*(14-15: 1), L**.** Algorithms. Robert Sedgewick. Addison-Wesley, 1983, viii + 551 pp, $31.95. [ISBN: 0-201-06672-6] A survey of some of the most important algorithms in computer science. Major sections cover mathematics algorithms, sorting, searching, string processing, geometric algorithms, and advanced topics. AO

**Computer Science, P.** Natural Language Communication with Pictorial Information Systems. Ed: Leonard Bolc. Symbolic Computation. Springer-Verlag, 1984, vii + 327 pp, $31. [ISBN: 0-387-13478-6] Reports of projects concerned with common representation and understanding of language and pictorial information. The papers discuss pictorial query languages and databases, natural language dialogue systems for scene analysis, interactive systems concerned with weather reporting, and interactive digital picture analysis. RM

**Computer Science, T(18), P.** Lecture Notes in Computer Science-166: STACS 84. Ed: M. Fontet, K. Mehlhorn. Springer-Verlag, 1984, vi + 338 pp, $15 (P). [ISBN: 0-387-12920-0] This monograph contains a collection of 27 papers presented at the Symposium on Theoretical Computer Science held in Paris, April, 1984. The papers address issues of current research in such theoretical areas as algorithmic analysis, Petri nets, formal languages, and information theory. MS

**Computer Science, P.** Lecture Notes in Computer Science-172: Automata, Languages and Programming. Ed: Jan Paredaens. Springer-Verlag, 1984, viii + 527 pp, $22 (P). [ISBN: 0-387-13345-3] Proceedings

of ICALP 84 (11th Colloquium on Automata, Languages, and Programming) covering the theoretical aspects of computer science. RM

**Computer Science, S*(15-17), L**.** Computing and Assessing Programming Languages: Ada, C, Pascal. Ed: A. Feuer, N. Gehani. Prentice-Hall, 1984, 271 pp, $24.95. [ISBN: 0-13-154857-3] A collection of papers discussing the programming languages Ada, C, and Pascal. The papers are grouped into sections containing comparisons, assessments, criticisms of these languages, and a final section on methodology for comparison and assessment of programming languages. AO

**Computer Science, T(17-18), P.** Supercomputers and Parallel Computation. Ed: D.J. Paddon. Institute of Math. & Its Applic. Conf. Ser., No. 1. Clarendon Pr, 1984, x + 258 pp, $39. [ISBN: 0-19-853601-1] This volume is based on the proceedings of a conference on parallel computing held at the University of Bristol in September, 1982. The book contains 18 research papers on the topics of VLSI parallel architecture, theory of parallel computing, vector and array processors, and algorithmic design for parallel machines. MS

**Computer Science, P*.** Information, Uncertainty, Complexity. J.F. Traub, G.W. Wasilkowski, H. Woźniakowski. Addison-Wesley, 1983, ix + 176 pp, $37.95. [ISBN: 0-201-07890-2] A research monograph which aims to create a general mathematical theory for understanding and dealing with uncertainty (as in approximate solutions to problems). The generality and power of the theory stems from the central role of information. A variety of application areas is presented to illustrate major concepts, together with an analysis of optimal or near-optimal algorithms. LCL

**Control Theory, P.** Optimal Control of Variational Inequalities. V. Barbu. Pitman, 1984, 298 pp, $23.95 (P). [ISBN: 0-273-08629-4] Monograph on first-order necessary conditions for optimality of (nonlinear) control problems governed by variational inequalities and semilinear equations of elliptic and parabolic types. Gives theoretical development of several simple problems (obstacle problem, Dirichlet problem). RM

**Control Theory, T(17-18), S, P.** Time Series Analysis, Identification and Adaptive Filtering. Daniel Graupe. Robert E Krieger, 1984, xv + 386 pp, $34.50. [ISBN: 0-88275-713-X] This book aims to provide tools for retrieving information in a time series when no prior knowledge of mathematical structure is available (i.e., "adaptive filtering of a-priori unknown data from a-priori unknown noisy background"). Assumes a background in probability and statistics, and linear control theory. Thirteen representative computer programs cover all major results of the text. LCL

**Applications, S(17), P.** Operational Calculus, Second Edition, V. I. Jan Mikusiński. Pergamon Pr, 1983, 321 pp, $27. [ISBN: 0-08-025071-8] Introduction to operational calculus and some applications, including: electric circuits, statics of beams, vibrations of a string, heat equation, and equations of telegraphy. Essentially the same as the first three parts of the first edition with an added chapter on applications to chromatography. BH

**Applications (Data Processing), T(15: 1).** Construction of Data Processing Software. John Elder. Ser. in Computer Sci. Prentice-Hall, 1984, xiv + 432 pp, $22.95 (P). [ISBN: 0-13-168675-5] Introduces commercial data processing problems. Assumes students have taken an introductory, Pascal-based programming course. Examines important concepts, techniques and algorithms using an extended Pascal in the first part of the book; the second part describes and utilizes ANSI Cobol. FA

**Applications (Economics), T(17-18: 2), S, P.** Advanced Econometric Methods. Thomas B. Fomby, R. Carter Hill, Stanley R. Johnson. Springer-Verlag, 1984, xix + 624 pp, $54. [ISBN: 0-387-90908-7] A comprehensive graduate-level introduction to its subject; one course in econometric methods is prerequisite. First section studies fundamental statistics of least-squares regression. Later sections treat violation of basic assumptions, special topics in linear regression, simultaneous equation models, frontiers. Each chapter ends with summary, guide to further reading, and exercises. PZ

**Applications (Economics), P.** Capital Flows and Exchange Rate Determination. Ed: Lawrence R. Klein, Wilhelm E. Krelle. Zeitschrift für Nationalökonomie Journal of Economics, Supp. 3. Springer-Verlag, 1983, ix + 220 pp, $41.10 (P). [ISBN: 0-387-81770-0] Seven papers presented at a September 1982 LINK meeting in Frankfurt, West Germany, published as a supplement to the Journal of Economics. Papers are equally divided between two-country and multi-country models. LAS

**Applications (Economics), T*(18: 1), S*, P.** Theory of Correspondences: Including Applications to Mathematical Economics. Erwin Klein, Anthony C. Thompson. Canadian Math. Soc. Ser. of Mono. & Adv. Texts. Wiley, 1984, xi + 256 pp, $39.95. [ISBN: 0-471-88016-7] Correspondences are graphs of relations; they model consumers' preferences. This monograph studies topologies on spaces of relations and spaces of subsets. Generalizes to correspondences the classical selection (E. Michael's) and fixed-point theorems (Brouwer's, Kakutani's). Introduces measure spaces to model economies; generalizes demand curves and Walras equilibria. Fun exercises in topology, measure theory, convex and functional analyses. Extremely well written. YN

**Applications (Engineering), P.** Mathematical Models for the Analysis and Optimization of Elastoplastic Structures. A.A. Cyras. Transl: L.W. Longdon. Ser. in Civil Engin. Halsted Pr, 1983, 121 pp, $39.95. [ISBN: 0-470-20020-0] Discrete models for predicting the behavior of a structure (deformation, etc.) at different stages of loading. LCL

**Applications (Engineering), P*.** Mechanics of Solids, Volume I: The Experimental Foundations of Solid Mechanics. James F. Bell. Springer-Verlag, 1984, xii + 813 pp, $36 (P). [ISBN: 0-387-13160-4] This book originally appeared in hardcover as Volume VIa/1 of Encyclopedia of Physics. JK

**Applications (Information Theory), P.** Transactions of the Ninth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. D Reidel, 1983, $50 each. Volume A, 332 pp [ISBN: 90-277-1499-1]; Volume B, 310 pp. [ISBN: 90-277-1500-9] Invited and contributed papers from a June 1982 conference in Prague. LAS

**Applications (Modelling), L.** Models of Reality: Shaping Thought and Action. Ed: Jacques Richardson. Lomond Books, 1984, xiv + 328 pp, $22.95. [ISBN: 0-912338-35-0] Essays on models as tools for problem-solving; cross-disciplinary. Many essays first appeared in the UNESCO journal Impact of Science on Society, Volume 31, No. 4. JRG

**Applications (Modelling), P.** Mathematical Modelling in Science and Technology. Ed: Xavier J.R. Avula, et al. Pergamon Pr, 1984, xvi + 1006 pp, $150. [ISBN: 0-08-030156-8] Typescript proceedings of the Fourth International Conference on Mathematical Modelling held at E.T.H. in Zurich, August 15-17, 1983. Seven invited lectures plus 150 papers cover a very wide range of contemporary modelling areas. LAS

**Applications (Physics), S(18), P.** Renormalization. Edward B. Manoukian. Pure & Appl. Math., V. 106. Academic Pr, 1983, xi + 204 pp, $37. [ISBN: 0-12-469450-0] First, proves the absolute convergence of Feynman integrals; uses tempered distributions and invokes Hironaka's results on resolution of singularities. Second, describes and proves a graph-theoretic algorithm to renormalize Feynman integrands. A mathematics book for physicists, not a physics book for mathematicians: never says how this mathematics is relevant to physics (see e.g., pp. 1-2 in Hwa and Toeplitz, Homology and Feynman Integrals, Benjamin, 1966). YN

**Applications (Physics), P.** Introduction to Quantum Statistical Mechanics. N.N. Bogolubov, N.N. Bogolubov, Jr. Transl: V.P. Gupta. World Scientific (US Dist: Heyden & Sons), 1982, viii + 299 pp, $19 (P); $34. [ISBN: 9971-950-31-6; 9971-950-04-9] In two sections: the first, described by the book's title, studies Hamiltonian systems in classical and quantum mechanics, thermodynamics. The second section is devoted to a new presentation of the second quantization method. Translated edition of a special lecture series delivered at the physics department of Moscow State University. PZ

**Applications (Physics), P.** Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables. A. Majda. Appl. Math. Sci., V. 53. Springer-Verlag, 1984, viii + 159 pp, $16.80 (P). [ISBN: 0-387-96037-6] From the book cover: "...systematic rigorous discussion of quasilinear hyperbolic systems in several space variables...emphasis on the special structure of the hyperbolic conservation laws of fluid dynamics and the interplay between theoretical results and these concrete equations." From material presented in topics courses and lectures at the University of California, Berkeley. JK

**Applications (Physics), T*(17-18: 1, 2), P.** General Relativity. Robert M. Wald. U of Chicago Pr, 1984, xiii + 491 pp, $30 (P); $50. [ISBN: 0-226-87033-2; 0-226-87032-4] Textbook on general relativity for graduate students from a modern, geometrical viewpoint with some recent advances and developments. Flexible. Suitable for one-term introductory course or for a full-year course. JK

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

# Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

## Oklahoma-Arkansas Section

The meeting of the Oklahoma-Arkansas Section was held on March 30-31, 1984 at Arkansas Tech University. The attendance was approximately 125 people.

Invited Addresses:

"Another Approach to Riemann-Stieltjes Integrals," by Kenneth A. Ross, University of Oregon.
"Transitions," by Jeanne Agnew, Oklahoma State University.

Short Presentations:

"Approximation of Periodic Functions by Trigonometric Sums," by Yuly Makovoz, Oklahoma State University.
"Small Factors of $2^N \pm 1$ for $N > 1200$," by Dale Woods, Central State University.
"Approximation of Abstract Analytic Functions on Intervals," by H.G. Burchard, Oklahoma State University.
"Quasi Circular Ground Track of Geosynchronous Satellites," by Philip C. Almes, East Central State University.
"On Semi-distributive $\mathcal{L}$-heredity Rings," by Carlos E. Caballero, Oklahoma State University.
"Ordered Sets Which Support Inverse Semigroups: Three Examples," by Lide Li, University of Arkansas.
"Isomorphisms of Incidence Rings," by Joel K. Haack, Oklahoma State University.
"Multiplicative Magic Squares," by Michelle Y. Penner, Oklahoma State University.
"Another Approach to Riemann-Stieltjes Integrals," by Kenneth A. Ross, University of Oregon.
"First and Second Moments of the Largest Part of a Partition," by Grey Williams, Hendrix College.
"Actuarial Problems," by Kathy Wyers, Arkansas Tech University.
"Tangents to the Parabola," by David Patocka, Oklahoma State University.
"Identities in Lie Rings and Lie Pseudo-rings," by Diane W. Corckett, Hendrix College.
"Maximizing Mathematics (Learning and Development) Through Review," by Sam Hartzler, Oklahoma City Public Schools.
"Using Calculus to Solve a Familiar Discrete Problem," by Jerry Johnson, Oklahoma State University.
"Teaching Experiential Applied Mathematics--TEAM," by James Choike and John Jobe, Oklahoma State University.
"Teaching Remedial College Algebra Using the Saxon Textbook," by Beverly Reed, Siloam Springs Public Schools.
"Transitions," by Jeanne Le Caine Agnew, Oklahoma State University.
"A Characterization of $T_1$ Compact Spaces," by Dragan Jankovic, University of Arkansas.
"Necessary and Sufficient Conditions for a Linear Function To Be Continuous," by Ray Hamlett, East Central University.
"Functions with $\alpha$-closed and Pre-closed Graphs," by Paul E. Long, University of Arkansas.
"Continuum Chainability and Monotone Decompositions in Certain Classes of Unicoherent Continua," by W. Dwayne Collins, Hendrix College.
"Cyclic Graphs," by Mahesh M. Hiremath, University of Arkansas.
"Degree One Maps of 3-manifolds," by Benny Evans, Oklahoma State University.
"Kuhn-Tucker Conditions in Nonlinear Mathematical Optimization," by Karen Anderson, Hendrix College.
"Maximization of Determinants in (1,0) Matrices," by David S. Rogers, Oklahoma Christian College.
"Commutativity and Distributivity: Different Perspectives," by Ben Marshall, Hendrix College.

## Northern California Section

The annual meeting of the Northern California Section was held on February 25, 1984 at San Francisco State University in San Francisco, California. There were 168 registered participants.

Invited Addresses:

"Indices of Power as a Measure of Centrality in Social Networks," by Guillermo Owen, University of Iowa.
"Combinatorial Problems with Surprising Solutions," by David Roselle, Virginia Polytechnic Institute.
"How Large an Elephant Will Fit in a Cube?" by Don Chakerian, University of California, Davis.
"The Mathematics of Perfect Shuffles," by Persi Diaconis, Stanford University.

Luncheon Speaker:

"Successful Education and Successful Life," by Peter Hilton, SUNY at Binghamton.

## Ohio Section

The annual fall meeting of the Ohio Section was held on November 2-3, 1984 at Muskingum College, New Concord, Ohio. Approximately 125 persons participated.

Invited Addresses:

"Renewing Undergraduate Mathematics," by Lynn Steen, St. Olaf College.
"Undecidable Mathematics," by Dennis Burke, Miami University.
"Teaching Experimental Applied Mathematics: The TEAM Project," by Phil Schmidt, University of Akron.

"Characterizing Trees by a Finite Sequence of Integers," by Lee Saunders, Miami University, Hamilton.
"Mathematics and the Law," by Mary Gray, American University.

Short Presentations:

"Computing a Metric on Dendograms," by Vic Norton, Bowling Green State University.
"Abundant Numbers on Odd Perfect Numbers," by Rich Laatsch, Miami University.
"Optimizing Boxes and Discs," by Tom Hern, Bowling Green State University.
"A Perspective on ICME-5," by Richard Shumway, Ohio State University.

Panel Discussion:

"The VisUMAP Project," by Zaven Karian (Moderator), Denison University; Lynn Steen, St. Olaf College; Wayne Carlson, Cranston-Csuri Productions.

## Metropolitan New York Section

The forty-third annual meeting of the Metropolitan New York Section was held at the College of Mt. Saint Vincent on May 6, 1984 with approximately 150 persons in attendance.

Invited Addresses:

"Increased Interaction in the Mathematics Classroom," by William F. Lucas, Cornell University.
"Sphere Packing in Higher Dimensions with Applications," by Neil J.A. Sloane, Bell Telephone Laboratories.

Short Presentations:

"Euler's 36 Officers Problem in Three Dimensions," by Joseph Arkin, New York Academy of Sciences; Paul Smith and E.G. Straus, University of California, Los Angeles.
"Teaching Basic to Liberal Arts Students on Pocket Computers Via Selected Topics in Finite Mathematics," by Joseph Ercolano, Baruch College, CUNY.
"A Simple Proof of Rodrigue's Formula Using the Weingarten Equations of Surface Theory," by Martin Lewinter, College at Purchase, SUNY.
"Clarification of Conservation of Liquid Quantity and Liquid Volume," by Egon Mermelstein, Academy of Aeronautics.
"Nonhomeomorphic, Equipotent Topological Spaces Possessing a Common Invariant," by Jay Schiffman, Kean College.
"Continuous Time Markov Chains Without Differential Equations," by Frederick Solomon, College at Purchase, SUNY.
"Covers, Piles, and Binary Search," by Aaron Todd, St. John's University.
"Good Numbers and the Multiplicative Property," by Clara Wajngurt, Queensborough Community College, CUNY.

Panel Discussion:

"Computers in Undergraduate Mathematics," by Raymond Greenwell, Hofstra University; Sheldon P. Gordon, Suffolk Community College; John Miller, City College.

## New Jersey Section

The fall meeting of the New Jersey section met at Rutgers University, New Brunswick, New Jersey on November 3, 1984. Approximately 60 people attended.

Invited Lectures:

"Classical Surface Theory in Minkowski 3-space," by Tilla Milnor, Rutgers University, New Brunswick.
"Reaching Math Anxious College Students--Some Psychological Approaches," by Irene Deitch, College of Staten Island, CUNY.
"What is Discrete Algorithmic Mathematics?" by Steven Maurer, Swarthmore College.

Special Presentation:

"A showing of videotapes and discussion of the materials from the M.A.A.'s TEAM project was presented by Ashby Foote, Rutgers University, New Brunswick.

## Seaway Section

The fall meeting of the Seaway Section was held on November 2-3, 1984 at St. Bonaventure University, St. Bonaventure, New York. There was a registered attendance of 91.

Invited Addresses:

"Gifted Students in Mathematics," by Robert Exner, Syracuse University.
"Computing and the Changing Nature of College Mathematics," by Donald Kreider, Dartmouth College.

Short Presentations:

"Upper Bound Graphs: A Polynomial--Time Algorithm," by Douglas Cashing, St. Bonaventure University.
"Determining Hidden Surfaces of Convex Polyhedra," by Paul O'Heron, Broome Community College.
"Teaching Experimental Applied Mathematics--Video Tape Project," by Edwin Hoefer, RIT and TEAM Leader for the Seaway Section.
"Solving Trigonometric Inequalities," by Boris Rakover, St. John Fisher College.
"International Congress of Mathematics-82," by Nura Turner, SUNY at Albany.
"Writing in the Math Classroom," by Marcia Birken, RIT.
"Appreciating Efficient Sorting," by Morton Goldberg, Broome Community College.
"Analytical Hierarchies and Decision Making," by Ronald Brzenk, Hartwick College.
"Polynomials, Map Coloring, and Flattening Equations," by Frank Bernhart, RIT.

## Northeastern Section

The spring meeting of the Northeastern Section was held jointly with the A.M.S. on June 29-July 1, 1984 at Plymouth State College in Plymouth, New Hampshire. There were approximately 110 registrants.

Invited Addresses:

"Teaching and Research in 'A Nation at Risk'," by Richard D. Anderson, Louisiana State University, Baton Rouge.
"Mathematics in Management Courses," by Alice T. Schafer, Simmons College, Wellesley College (Emeritus).
"Accreditation and Certification of MA and CS Programs," by John Dalphin, Norwich University.
"The Fourth Dimension," by Stephanie Troyer, University of Hartford.

Workshops:

"Applications of Undergraduate Mathematics," by Solomon A. Garfunkel, Executive Director, COMAP.
"UNIX," by Roger Kleinpeter and Michael Pearson, Plymouth State College.
"MACSYMA and SMP," by Ken Lane, Colby College.
"Karel the Robot: A Gentle Microworld of Structured Programming," by William J. Taffe, Plymouth State College.

Panel Discussion:

"The Crisis in Mathematics Teachers Training," by Fernand Prevost (Moderator), New Hampshire State Department of Education; Joan Mundy, University of New Hampshire; Robert Rosenbaum, Wesleyan University; Martin Badoian, Canton High School, Canton, Massachusetts.

## North Central Section

The fall meeting of the North Central Section was held at Moorhead State University, Moorhead, Minnesota on October 26-27, 1984. There were 103 registrants.

Invited Addresses:

* "The University of Minnesota Talented Youth Project," by Harvey Keynes, University of Minnesota.
"Some Unexpected Results in Elementary Mathematics," by Ivan Niven, University of Oregon.

Short Presentations:

* "Composition for Two Peanos," by Paul Fjelstad, St. Olaf College.
* "Time in Special Relativity," by Abraham Ungar, North Dakota State University.
"Mathematical Art by Computer," by Douglas Dunham, University of Minnesota, Duluth.
* "Proteomorphisms--On Generalizing Duality and Similarity," by Thomas Sibley, St. John's University.
* "Continued Roots," by Walter S. Sizer, Moorhead State University.
* "An Extended Hatzenbuhler-Mattson Technique," by Marlon C. Rayburn, University of Manitoba.
* "Contraction Mappings and Approximation of Roots," by Michael Tangredi, College of St. Benedict.
* "Some Properties of Pascal Triangle Matrices," by Dan Kalman, Augustana College.

Special Sessions:

Special sessions were held on "Policies for Secondary Mathematics Education," and "On Problem Solving."

9. M. Lerch, Dalši studie v oboru Malmsténovských řad, Rozpravy Česke Akad. 3, no. 28, 1894, 63 pp.
10. C. J. Moreno, The Chowla-Selberg formula, J. Number Theory, 17 (1983) 226–245.
11. T. Shintani, On special values of zeta-functions of totally real quadratic fields, Proc. Inter. Congr. Math., Helsinki, 1978, vol. 2, Acad. Sci. Fennica, Helsinki, 1980, pp. 591–597.
12. E. C. Titchmarsh, The Theory of the Riemann Zeta-Function, Clarendon Press, Oxford, 1951.
13. A. Weil, Elliptic Functions According to Eisenstein and Kronecker, Springer-Verlag, Berlin, 1976.
14. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed., University Press, Cambridge, 1962.

---

## 138. MISCELLANEA

Just as surely as our understanding of Nature is really valid only to the extent that it is mathematical, so also our understanding of higher domains must be based on mathematical models.

—R. Steiner (1861–1925), quoted in *Mathematical Reviews* 82k : 51001.

Each natural science contains genuine Science only to the extent that it contains Mathematics.... It may be that a formal philosophy of Nature in general, that is, a philosophy that deals only with general concepts, is possible without Mathematics, but a formal natural science dealing with definite objects (whether Physics or Psychology) is possible only with the use of Mathematics, and since each natural science contains only as much genuine Science as it contains *a priori* knowledge, it follows that a natural science is a genuine Science only to the extent that Mathematics can be applied to it.

—Immanuel Kant, Metaphysische Anfangsgründe der Naturwissenschaft, 1876.

*Perhaps we should sometimes listen to the other side. Goethe, for example, did not like Mathematics.*

Mathematicians are amazing people. In virtue of their accomplishments, they have set themselves up as a universal guild and will acknowledge only what suits their circle, what their method of organization can produce. A prominent mathematician once said, when someone strongly recommended a topic in Physics, "But can't it be reduced to calculation?"

—Goethe, Maximen und Reflexionen, no. 1277

It surely does not follow that the hunter who kills the game must also cook it. A cook might go hunting and shoot well; but he would be badly mistaken if he claimed that only a cook can be a good shot. It seems to me that this is the situation of mathematicians who claim that nobody can understand or discover physical phenomena without being a mathematician, since they should be pleased enough if the meat is brought to their kitchen for them to lard it with formulas and dress it as they like.

—Goethe, Maximen und Reflexionen, no. 1280

---

### ANSWER TO PHOTO ON PAGE 93

The photo is of Louis de Branges, who proved in 1984 the conjecture Bieberbach made in 1916, by establishing a strong inequality proposed by the Russian function-theorist I. M. Milin.

## ON HALLEY'S ITERATION METHOD

WALTER GANDER
*Seminar für angewandte Mathematik ETH, CH-8092 Zürich, Switzerland*

**1. Introduction.** Solving a nonlinear equation is a problem that has occupied mathematicians for many centuries and many numerical analysts of the present generation first learned how to do it from P. Henrici's textbook [7] which has now been replaced by [**8**].

Nearly 300 years ago in 1694 Edmund Halley published a paper [**6**] in Latin where he presents a new method to compute roots of a polynomial. Halley is well known for first computing the orbit of the Halley comet, which he observed in 1682 and which will soon visit us again in 1986.

Halley generalized an iteration formula due to Lagney for computing the cubic root of a number and obtained an iteration to compute roots of a polynomial. Halley did not use calculus to derive his formula. If we try to generalize and translate his derivation into modern mathematical notation we can argue as follows:

Let $f$ be a $C^2$ function and $x_k$ an approximation of a zero $s$ of $f$. We replace the equation $f(x) = 0$ by $T(h) = 0$, where $T$ is the Taylor polynomial of degree 2 of $f$

$$T(h) = f(x_k) + f'(x_k)h + \frac{f''(x_k)}{2}h^2 \tag{1}$$

and $h = x - x_k$. Now if $x_k$ is a good approximation of $s$, then neglecting $h^2$ we get from

$$T(h) = 0 \tag{2}$$

the *Newton correction*

$$h = -f(x_k)/f'(x_k). \tag{3}$$

Since we neglected $h^2$, the denominator in (3) is wrong: equation (3) should be

$$h = -f(x_k)\Big/\left(f'(x_k) + \frac{f''(x_k)}{2}h\right). \tag{4}$$

Therefore, replacing $h$ in the denominator of (4) by the Newton correction (3), we obtain the *Halley correction*

$$h = \frac{-f(x_k)}{f'(x_k) - \dfrac{f''(x_k)f(x_k)}{2f'(x_k)}}. \tag{5}$$

For the following discussion we rearrange Halley's iteration formula (5) to

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \cdot \frac{1}{1 - \dfrac{t(x_k)}{2}}, \tag{6a}$$

where

$$t(x) = \frac{f''(x)f(x)}{f'(x)^2}. \tag{6b}$$

In his paper [**6**], Halley calls the iteration formula (6a) the *rational formula*. He also considers the *Euler correction*

$$h = \frac{-f'(x_k) \pm \sqrt{f'(x_k)^2 - 2f(x_k)f''(x_k)}}{f''(x_k)} \tag{7}$$

that is obtained by solving the quadratic equation (2) and which he calls the *irrational formula*. If we rearrange (7) and compute the correction closest to zero, we get Euler's iteration

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \cdot \frac{2}{1 + \sqrt{1 - 2t(x_k)}}, \tag{8}$$

where $t$ is defined by (6b).

In the following we shall show how Halley's method and also many other third order iteration formulas may be derived algebraically using an elementary technique. For geometric interpretations of Halley's method we refer to [**1**], [**2**].

**2. Algebraic interpretation.** Halley's method (6) belongs to the class of *one point iteration methods without memory* [**12**]

$$x_{k+1} = F(x_k). \tag{9}$$

We consider the special case where the iteration function $F$ has the form

$$F(x) = x - \frac{f(x)}{f'(x)} G(x). \tag{10}$$

For the following discussion we assume that $s$ is a simple zero of $f$ and that $f$ and $G$ have a sufficient number of continuous derivatives in a neighborhood of $s$. It is well known [**11**] that the iteration (9) is of second order if

$$F'(s) = 0, F''(s) \neq 0. \tag{11}$$

Let $u(x) = f(x)/f'(x)$. Then we have $u'(x) = 1 - t(x)$ where $t$ is defined by (6b). It follows that

$$u(s) = 0 \text{ and } u'(s) = 1. \tag{12}$$

Differentiating (10) we get

$$F'(x) = 1 - u'(x)G(x) - u(x)G'(x). \tag{13}$$

And using (12) we get $F'(s) = 1 - G(s)$.

LEMMA 1. *The iteration* (9) *with F defined by* (10) *is of second order if and only if* $G(s) = 1$.

We note that for the special case $G(x) \equiv 1$ we have Newton's method for the equation $f(x) = 0$.

The iteration (9) is of third order if

$$F'(s) = F''(s) = 0, F'''(s) \neq 0. \tag{14}$$

Differentiating (13) we get

$$F''(x) = -u''(x)G(x) - 2u'(x)G'(x) - u(x)G''(x).$$

Since

$$u''(x) = -t'(x) = -\frac{f''(x)}{f'(x)} + f(x)\left(\frac{2f''(x)^2}{f'(x)^3} - \frac{f'''(x)}{f'(x)^2}\right),$$

it follows that

$$u''(s) = -\frac{f''(s)}{f'(s)} = -t'(s) \tag{15}$$

and

$$F''(s) = \frac{f''(s)}{f'(s)} G(s) - 2G'(s).$$

LEMMA 2. *The iteration* (9) *with $F$ defined by* (10) *is of third order if and only if* $G(s) = 1$ *and* $G'(s) = (1/2)f''(s)/f'(s)$.

The assumptions of Lemma 2 are not helpful for choosing a function $G$ since they need the knowledge of the zero $s$. However, if we choose

$$G(x) = H(t(x)) \tag{16}$$

with some function $H$, then we have

$$G(s) = H(0)$$

and

$$G'(s) = H'(0)t'(s).$$

Therefore the assumptions of Lemma 2 are simply $H(0) = 1$ and $H'(0) = 1/2$. We get the following theorem.

THEOREM. *Let $s$ be a simple zero of $f$ and $H$ any function with $H(0) = 1$, $H'(0) = 1/2$ and $|H''(0)| < \infty$. The iteration $x_{n+1} = F(x_n)$ with*

$$F(x) = x - \frac{f(x)}{f'(x)} H(t(x))$$

*where*

$$t(x) = \frac{f(x)f''(x)}{f'(x)^2}$$

*is of third order.*

Many well-known third order iterative methods are special cases of this theorem:

(1) Halley's method

$$H(t) = \left(1 - \tfrac{1}{2}t\right)^{-1} = 1 + \frac{1}{2}t + \frac{1}{4}t^2 + \cdots.$$

(2) Euler's formula

$$H(t) = 2\left(1 + \sqrt{1 - 2t}\right)^{-1} = 1 + \frac{1}{2}t + \frac{1}{2}t^2 + \cdots.$$

(3) Hansen-Patrick family [**5**]

$$H(t) = (a + 1)\left(a + \sqrt{1 - (a + 1)t}\right)^{-1} = 1 + \frac{1}{2}t + \frac{a + 3}{8}t^2 + \cdots.$$

(4) Ostrowski's square root iteration [**10**]

$$H(t) = (1 - t)^{-0.5} = 1 + \frac{1}{2}t + \frac{3}{8}t^2 + \cdots.$$

(5) Quadratic inverse interpolation

$$H(t) = 1 + \frac{1}{2}t.$$

Not all the third order iteration methods are special cases of this theorem. It is possible to describe all methods that do not explicitly depend on $s$ by using the Schröder iteration functions [**9**, p. 531]. One can show [**4**] that all third order methods are given by the iteration function (10) with

$$G(x) = H(t(x)) + f(x)^2 b(x),$$

where $b$ is an arbitrary function which is bounded for $x \to s$.

**References**

**1.** G. Alefeld, On the convergence of Halley's method, this MONTHLY, 88 (1981) 530–536.

2. G. H. Brown, Jr., On Halley's variation of Newton's method, this MONTHLY, 84 (1977) 726–728.
3. M. Davis and B. Dawson, On global convergence of Halley's iteration formula, Numer. Math., 24 (1975) 133–135.
4. H. Ehrmann, Konstruktion und Durchführung von Iterationsverfahren höherer Ordnung, Arch. Rational Mech. Anal., 4 (1959) 65–88.
5. E. Hansen and M. Patrick, A family of root finding methods, Numer. Math., 27 (1977) 257–269.
6. E. Halley, Methodus Nova, Accurata & Facilis Inveniendi Radices Aequationum Quarumcumque Generaliter, Sine Praevia Reductione, Philos. Trans. Roy. Soc. London, 18 (1694) 136–148.
7. P. Henrici, Elements of Numerical Analysis, Wiley, New York, 1964.
8. ______, Essentials of Numerical Analysis, Wiley, New York, 1982.
9. ______, Applied and Computational Complex Analysis, Wiley, New York, 1974.
10. A. M. Ostrowski, Solution of Equations and Systems of Equations, Academic Press, New York, 1973.
11. J. Stoer and R. Bulirsch, Introduction to Numerical Analysis, Springer-Verlag, New York, 1980.
12. J. F. Traub, Iterative Methods for Solution of Equations, Prentice-Hall, 1964.

## A CONVERSE FOR THE CAYLEY-HAMILTON THEOREM

CARMEN CHICONE, N.J. KALTON*, AND IRA J. PAPICK
*Department of Mathematics, University of Missouri-Columbia, Columbia, MO 65211*

The Cayley-Hamilton theorem asserts that every $n \times n$ matrix $A$ satisfies its characteristic polynomial, $\det(\lambda I - A)$. This note deals with the problem of characterizing those polynomials for which the Cayley-Hamilton theorem holds. Informally stated our result is that the polynomials which a square matrix satisfies are precisely the multiples (in a ring of polynomials) of the characteristic polynomial.

Some notation is necessary to make a precise statement of our theorem. If $X$ denotes the $n \times n$ matrix of indeterminants $(x_{ij})$, then it is apparent that $\det(X)$ is a polynomial in $n^2$-variables. Moreover, if $F(x_{ij})$ is any polynomial in $n^2$-variables with coefficients in a commutative ring $R$ with identity, then by $F(X)$ we shall mean $F(x_{ij})$. We now restate our problem: Characterize those polynomials $F(X)$ having the property that every $n \times n$ matrix $A$ with entries in $R$ satisfies the polynomial $F(\lambda I - A)$. We shall call such polynomials $F(X)$ Cayley-Hamilton polynomials, and we may now state a precise converse of the Cayley-Hamilton theorem.

THEOREM 1. *Let $R$ be an infinite (commutative) integral domain, and let $F(X)$ be a polynomial in $n^2$-variables with coefficients in $R$. Then $F(X)$ is a Cayley-Hamilton polynomial if and only if $F(X) = \det(X)G(X)$, where $G(X)$ is a polynomial in $n^2$-variables with coefficients in $R$.*

Before proceeding we make two observations. The "if" direction of the theorem is clear. For if $A$ is an $n \times n$ matrix with entries in $R$, and if $F(X) = \det(X)G(X)$, then $F(\lambda I - A) = \det(\lambda I - A)G(\lambda I - A)$. Hence, by the Cayley-Hamilton theorem, $A$ satisfies the polynomial $F(\lambda I - A)$. Secondly, the theorem is false without some assumptions on $R$. For example if $R = Z_2$, the field with two elements, define $F(X) = (x_{12} + x_{21})x_{11}x_{22}$. Then given

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

with $a_{ij} \in Z_2$, we see that

$$F(\lambda I - A) = (a_{12} + a_{21})(\lambda - a_{11})(\lambda - a_{22}).$$

Thus $F(\lambda I - A) \equiv 0$ unless $A$ is either upper or lower triangular and not diagonal, and in this case

$$F(\lambda I - A) = (\lambda - a_{11})(\lambda - a_{22}) = \det(\lambda I - A).$$

*Supported in part by NSF grant MCS-8001852.

follows from the equation

$$F(X,X)=Q(X,X)D(X,X)+R(X,X),$$

since by the Cayley-Hamilton theorem $D(X,X)=0$ and, as shown above, $F(X,X)=0$. Thus if

$$R(X,\lambda)=u_r(X)\lambda^r+\cdots+u_0(X),$$

we have $u_r(X)X^r+\cdots+u_0(X)I=0$.

We now claim that the characteristic polynomial, $D(X,\lambda)\in S[\lambda]$ of $X$ has $n$ distinct roots (in some splitting field of the field of fractions of $S$). Once this is shown, the proof will be complete, for the minimal polynomial of the matrix $X$ will be degree $n$ and hence we can conclude that $u_r(X)=\cdots=u_0(X)=0$, i.e., $R(X,\lambda)=0$.

To establish our claim we note that the *discriminant* $\Delta(X)$ of $D(X,\lambda)$ is a homogeneous polynomial in the coefficients of $D(X,\lambda)$, and $\Delta(X)=0$ if and only if $D(X,\lambda)$ has a repeated root [**3**, p. 288]. If $A\in M_n(R)$, then $\Delta(A)$ is the discriminant of $D(A,\lambda)$; hence by selecting $A$ to be diagonal with $n$ distinct eigenvalues ($R$ is infinite!), we see that $\Delta(X)$ cannot vanish identically. This completes the proof.

Now to derive the theorem from Lemma 2. Let $G(X)$ be a Cayley-Hamilton polynomial, and define $F(X,\lambda)=G(\lambda I-X)$. Then $F(A,A)=0$ for each $A\in M_n(R)$, and thus it follows from Lemma 2 that $D(X,\lambda)$ is a factor of $F(X,\lambda)$, i.e., $F(X,\lambda)=D(X,\lambda)Q(X,\lambda)$. Hence, putting $\lambda=0$ and replacing $X$ by $-X$ we have

$$G(X)=\det(X)Q(X,0),$$

where

$$Q(X,0)\in R[X].$$

**References**


1. W. Fulton, Algebraic Curves, W. A. Benjamin, New York, 1969.
2. T. Hungerford, Algebra, Holt, Rinehart and Winston, New York, 1974.
3. J. Uspensky, Theory of Equations, McGraw-Hill, New York, 1948.


## AN ELEMENTARY CHARACTERIZATION OF WEAK CONVERGENCE OF MEASURES

JERRY JOHNSON
*Department of Mathematics, Oklahoma State University, Stillwater, OK 74078*

In 1977, Högnäs [**2**] solved a problem of P. Lévy which, roughly put, asks for conditions on a family $\{K_s\}$ of functions so that for each continuous function $f$ on $[0,1]$ we have continuity of $s\to\int_0^1 f\,dK_s$. The solution follows immediately from his main theorem, whose proof is fairly long. This note presents a short, elementary proof of his theorem, using only a few basic results from a standard real variables text such as Royden [**3**].

Our notation is as follows: $C$ denotes the continuous functions on $[0,1]$ with supremum norm $\|\cdot\|_\infty$. If $f$ and $K$ are functions on $[0,1]$, $\int f\,dK$ will denote the Riemann-Stieltjes integral when it exists (see [**1**] for the definition and basic facts). $\lambda$ will denote Lebesgue measure on $[0,1]$.

The following five items used in the proof are stated in the *least* possible generality, reflecting the assumptions in the theorem.

(1) The Arzela-Ascoli Theorem: An equicontinuous bounded subset of $C$ is relatively compact.
(2) The Stone-Weierstrass Theorem: The polynomials are dense in $C$.
(3) If $\mu$ is a signed measure on $[0,1]$, $f\in C$ and $K(x)=\mu[0,x]$ for $0\leqslant x\leqslant 1$, then $K$ is of bounded variation and $\int f\,d\mu=K(0)f(0)+\int f\,dK$.
(4) Integration by parts: If either $\int f\,dK$ or $\int K\,df$ exists, so does the other and $\int f\,dK+\int K\,df=Kf|_0^1$ (see [**1**, p. 195]).

(5) If $f$ is absolutely continuous and $K$ is of bounded variation, then $\int K\,df = \int Kf'd\lambda$. (Suppose $\phi$ is bounded and measurable. Let $f(x) = \int_0^x \phi\, d\lambda$ and replace $f'$ by $\phi$. This is the context in which (5) is used and it is an exercise to prove this special case.)

We are now ready to present the main theorem in [**2**].

THEOREM [**2**, p. 182]. *Let $\{\mu_\alpha\}$ be a bounded net (or sequence) of signed Borel measures on* $[0,1]$; *i.e., there is a number $M > 0$ such that $|\int f d\mu_\alpha| \leqslant M\|f\|_\infty$ for $f \in C$. Define $K_\alpha(x) = \mu_\alpha[0,x]$ for $0 \leqslant x \leqslant 1$. Then the following are equivalent*:

(A) $\lim_\alpha \int f d\mu_\alpha = 0$ *for each $f \in C$ (i.e., $\{\mu_\alpha\}$ converges weak* to zero).*
(B) $\lim_\alpha(\int|K_\alpha|d\lambda + |K_\alpha(1)|) = 0$.

*Proof*: Assume (B). Let $f$ be continuously differentiable on $[0,1]$. Then

$$\begin{aligned}\left|\int f d\mu_\alpha\right| &= \left|\int f dK_\alpha + f(0)K_\alpha(0)\right|\\ &= \left|fK_\alpha\Big|_0^1 - \int K_\alpha\, df + f(0)K_\alpha(0)\right|\\ &\leqslant |f(1)K_\alpha(1)| + \left|\int K_\alpha\, df\right|\\ &\leqslant |f(1)K_\alpha(1)| + \int |K_\alpha f'|\, d\lambda\\ &\leqslant |f(1)K_\alpha(1)| + \|f'\|_\infty \int |K_\alpha|\, d\lambda\end{aligned}$$

which converges to zero. Hence $\lim_\alpha \int f d\mu_\alpha = 0$ for $f$ in a dense subset of $C$. Since $\{\mu_\alpha\}$ is bounded, it is an easy exercise to show that $\lim_\alpha \int f d\mu_\alpha = 0$ for all $f \in C$.

Next assume (A). First $K_\alpha(1) = \int 1\, d\mu_\alpha$, which converges to zero. Now, for each $\alpha$, choose a function $\phi_\alpha$ so that $\phi_\alpha$ is measurable, $|\phi_\alpha(t)| = 1$ for all $t$ and $\phi_\alpha K_\alpha = |K_\alpha|$. Define $f_\alpha(x) = \int_0^x \phi_\alpha\, d\lambda$ for $0 \leqslant x \leqslant 1$. It is easy to check that $\{f_\alpha\}$ is an equicontinuous bounded set in $C$ and hence is relatively compact in $C$. Now, it is another easy exercise to show that since $\{\mu_\alpha\}$ is bounded and $\lim_\alpha \int f d\mu_\alpha = 0$ for $f \in C$, then $\lim_\alpha \int f d\mu_\alpha = 0$ uniformly for $f$ in a compact set. From this, it follows that $\lim_\alpha \int f_\alpha\, d\mu_\alpha = 0$. Finally,

$$\begin{aligned}\int f_\alpha\, d\mu_\alpha &= \int f_\alpha\, dK_\alpha + f_\alpha(0)K_\alpha(0)\\ &= f_\alpha K_\alpha|_0^1 - \int K_\alpha\, df_\alpha + f_\alpha(0)K_\alpha(0)\\ &= K_\alpha(1)f_\alpha(1) - \int K_\alpha f_\alpha'\, d\lambda\\ &= K_\alpha(1)f_\alpha(1) - \int K_\alpha\phi_\alpha\, d\lambda\\ &= K_\alpha(1)f(1) - \int |K_\alpha|\, d\lambda.\end{aligned}$$

Therefore, $\int|K_\alpha|\,d\lambda = K_\alpha(1)f_\alpha(1) - \int f_\alpha\, d\mu_\alpha$, which converges to zero since $\{f_\alpha(1)\}$ is bounded. This completes the proof.

## References


**1.** T. M. Apostol, Mathematical Analysis, Addison-Wesley, Reading, MA, 1960.

**2.** Göran Högnas, Characterization of weak convergence of signed measures on [0,1], Math. Scand., 41 (1977) 175–184.

**3.** H. L. Royden, Real Analysis, Macmillan, New York, NY, 1968.

# INTERLACED SECOND CATEGORY SETS

HARVEY DIAMOND AND GREGORY GELLES
*Department of Mathematics, West Virginia University, Morgantown, WV 26506*

In this note we construct a pair of disjoint subsets of the real line, each of which intersects every interval in a set of second category. Our intuitive motivation is the creation of disjoint subsets each having a "large" representation in every interval of the real line, where "large" is interpreted in a topological way as "second category." In a measure-theoretic sense "large" may be interpreted as nonnull (i.e., positive outer measure) or as positive measure; in either case, well-known examples exist of disjoint sets which are measure-theoretically "large" on every interval. A nonmeasurable example may be found in [**3**, p. 70], and a recent measurable example is given in [**9**]. Our intention here is a simple construction of analogous examples for second category largeness. We will provide both measurable and nonmeasurable examples.

The existence of analogies between measure and category have long been the subject of investigation. For example, a formal link between second category and positive outer measure can be found in the Sierpinski-Erdős Duality Theorem [**6**]:

> *Let $P$ be any proposition involving solely the notions of measure zero, first category, and notions of pure set theory. Let $P^*$ be the proposition obtained from $P$ by interchanging the terms "nullset" and "set of first category" wherever they appear. Then each of the propositions $P$ and $P^*$ implies the other, assuming the continuum hypothesis.*

A further connection that our second category sets have with an established area of mathematics concerns the Property of Baire. Briefly, the collection of sets having the Property of Baire is the smallest $\sigma$-algebra containing the open sets and the sets of first category. In [**4**, p. 88] it is proved that a set of real numbers does not have the Property of Baire if and only if there exists an interval such that the set and its complement both have second category intersection with every subinterval of that interval. Thus, the disjoint sets that concern us may be equivalently characterized as sets whose intersection with any interval does not have the Property of Baire. A fuller discussion of the Property of Baire may be found in [**4**] and [**6**].

We begin our construction by noting that there already exist examples of disjoint nonmeasurable sets which are second category on every interval. (This property does not appear to have been previously observed.) For instance, the classic Vitali construction of a nonmeasurable set as in [**8**, p. 52] provides such an example. This construction produces a set $E$ such that $\{E + r : r \in Q\}$ is a countable partition of $R$, where $Q$ is the set of rational numbers. Since $R$ is of second category, $E + r$ must be of second category for all $r$. Now suppose that $I$ is an interval of length $2^{-k}$, $k \in N$, where $N$ is the set of positive integers. Then one of the sets $(E + n2^{-k}) \cap I$, where $n$ ranges over the set of integers $Z$, must be of second category. Consequently, the set $A \equiv \cup\{E + n2^{-k} : n \in Z, k \in N\}$ meets every interval in a set of second category, as does the (disjoint) translate $A + 1/3$.

The construction of two disjoint *measurable* sets which meet every interval in a set of second category can now be effected by intersecting the disjoint sets $A$ and $A + 1/3$ above with any measure zero set having a first category complement. (Such measure zero sets are well known. An example is constructed in [**2**, p. 99].) Indeed, if $V$ has first category complement, then $A \cap V$ differs from $A$ by a set of first category and thus retains the desired property.

Finally we present a different, more explicit construction of a measurable example, which is interesting in that it reveals some of the unusual properties of the set of nonnormal numbers and parallels the Vitali construction of a nonmeasurable set. (Recall that a number is normal to base $r$ ($r$ an integer larger than 1) if in its base $r$ expansion all blocks of digits with the same length occur with equal frequency. Thus for instance, each digit occurs with frequence $1/r$. A more rigorous definition may be found in [**5**, p. 95]. A number that is not normal to base $r$ is said to be

nonnormal to base $r$.) The construction proceeds by obtaining a countable partition by translations of a second category set of measure zero. In contrast to the example in [**2**, p. 99], our purposes require a measure zero, second category set whose members can be characterized specifically, for instance, by their base two expansions. The set of numbers nonnormal to base two is such a set.

A proof that the set of nonnormal numbers to any base $r$ has measure zero can be found in [**5**, p. 103]. A different proof using probabilistic arguments is outlined in [**1**, pp. 105–107]. The proof uses the law of large numbers and the fact that Lebesgue measure can be obtained as the probability distribution of real numbers generated at random by choosing each digit independently with each of the $r$ possibilities equally likely.

The fact that the numbers nonnormal to base $r$ form a set of second category (in fact have first category complement) does not seem to be widely known. Proofs can be found in [**6**] and [**7**], but we will furnish a short argument for the case of base two.

In the base two expansion of a number $x$, let $L(k, x)$ denote the number of zeroes that appear in the first $k$ places after the decimal point. Now define for each $m \in N$ the union of intervals

$$E_m = \{x : x \in (n2^{-m}, n2^{-m} + 2^{-3m}), n \in Z\}.$$

Note that if $x \in E_m$, then the base two expansion of $x$ has zeroes in places $m+1$ through $3m$ so that $L(3m, x) \geqslant 2m$. This means in particular that the set $\overline{\lim} E_m = \{x : x \in E_m$ for infinitely many $m\}$ consists entirely of numbers nonnormal to base two. Set theoretically,

$$\overline{\lim}\, E_m = \bigcap_{M=1}^{\infty} \bigcup_{m=M}^{\infty} E_m.$$

The set $\cup_{m=M}^{\infty} E_m$ is open and contains the dense set of points $\{n2^{-m} + 2^{-3m-1}\}, n \in Z, m \geqslant M$ so its complement is nowhere dense. Therefore the complement of $\overline{\lim} E_m$ is the countable union of nowhere dense sets and thus of first category. Therefore the set of numbers nonnormal to base two has a first category complement and so is a set of second category.

The next step is the partition of the set of numbers nonnormal to base two, which we denote by $B$. Let $D$ denote the set of dyadic rationals, where each element of $D$ has a base two expansion which terminates in an infinite string of zeroes. We observe that if $x \in B$ and $d \in D$, then $x + d \in B$. Divide $B$ into equivalence classes according to $x \sim y$ if and only if $x, y \in B$ and $x - y \in D$. Using the Axion of Choice, form a set $F$ consisting of one element from each equivalence class. Clearly, $B = \cup_{d \in D}(F + d)$. Since $B$ is of second category, $F + d$ must be of second category for all $d \in D$. Let $D_e$ denote the set of dyadics having an even number of ones to the right of the decimal point in its base two expansion. Define $D_0$ similarly for odd numbers of ones ($D_0 = D \setminus D_e$). Note that $D_e$ and $D_0$ are both dense in $R$. Define the two sets

$$B_1 = \bigcup_{d_e \in D_e} (F + d_e) \text{ and } B_2 = \bigcup_{d_0 \in D_0} (F + d_0).$$

We show that $B_1$ meets every interval in a set of second category. Indeed, if $I$ is any interval, then the density of $D_e$ implies that

$$F = \bigcup_{d_e \in D_e} \{[(F + d_e) \cap I] - d_e\}.$$

Since $F$ is of second category, one of the sets $(F + d_e) \cap I$ must be of second category and thus $B_1 \cap I$ is of second category. A similar argument applies to $B_2$.

## References


1. K. L. Chung, A Course in Probability Theory, Academic Press, New York, 1974.
2. B. Gelbaum and J. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.
3. P. Halmos, Measure Theory, Springer-Verlag, New York, 1974.
4. K. Kuratowski, Topology, Academic Press, New York, 1966.

5. I. Niven, Irrational Numbers, Mathematical Association of America, Washington, DC, 1956.
6. J. Oxtoby, Measure and Category, Springer-Verlag, New York, 1971.
7. J. Oxtoby and S. Ulam, Measure-preserving homeomorphisms and metrical transitivity, Ann. of Math., 42 (2) (1941) 874–920.
8. W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1966.
9. ______, Well-distributed measurable sets, this MONTHLY, 90 (1) (1983) 41–42.

## LION-HUNTING WITH LOGIC*

HOUSTON EULER
*c/o Leon Harkleroad, 4417 Woodgate Lane, Louisville, KY 40220*

Over the years there has developed a body of literature on the use of mathematical techniques to catch lions ([**1**]–[**6**]). In this literature there has been a comparative shortage of proofs based on mathematical logic. If those of us who commit logic believe in the vitality of our field, we cannot afford to allow such a shortage to continue. The following proofs, then, are offered as a first step towards rectifying the situation.

1. *Nonstandard Analysis.* In a nonstandard universe (namely, the land of Oz [7]), lions are cowardly and may be caught easily. By the transfer principle, this likewise holds in our (standard) universe.

2. *Set Theory.* If the set of lions is bounded, you can simply build a cage around the boundary. So assume that the set of lions is unbounded. It will then have an element in common with a stationary set. But a stationary lion is trivial to capture.

3. *Set Theory.* Assume $V = L$. Since the lion is in the universe, it is constructible. So just carry out its construction within a cage in the first place.

4. *Set Theory.* Assume AC. Perform a Tarski-Banach decomposition on the lion to halve its size. Repeat until the lion is small enough to be captured easily.

5. *Recursion Theory.* Assume you can capture a lion. Having done so, you can easily bring it to a standstill, and you would thus have a solution to the halting problem. Since the halting problem is unsolvable, you *cannot* capture a lion after all.

In conjunction with the previous results, we have

COROLLARY. *Mathematics is inconsistent.*

This corollary, besides being of intrinsic interest, also provides solutions to the Riemann Hypothesis, Fermat's Last Theorem, and other questions (besides giving a proof of the Four-Color Theorem that does not require a computer!).

### References

1. H. Pétard, A contribution to the mathematical theory of big game hunting, this MONTHLY, 45 (1938) 446–7.
2. I. J. Good, A new method of catching a lion, this MONTHLY, 72 (1965) 436.
3. C. Roselius, On a theorem of H. Pétard, this MONTHLY, 74 (1967) 838–9.
4. O. Morphy, Some modern mathematical methods in the theory of lion hunting, this MONTHLY, 75 (1968) 185–7.
5. P. L. Dudley, G. T. Evans, K. D. Hansen, I. D. Richardson, Further techniques in the theory of big game hunting, this MONTHLY, 75 (1968) 896–7.
6. J. Barrington, 15 new ways to catch a lion, Manifold, Issue 18 (1976), reprinted in Seven Years of Manifold, Shiva Publishing, 1981.
7. L. F. Baum, The Wonderful Wizard of Oz, 1900.

*This research received compact support from Tombs grant LWH-42755.

## AN INGRAINED ERROR CONCERNING RESULTANTS

Ernst Snapper
*Department of Mathematics, Dartmouth College, Hanover, NH 03755*

All books I know in modern algebra which discuss the resultant of two polynomials, including the famous books by Jacobson [1], Lang [2] and van der Waerden [3], state the following wrong theorem.

Let $F$ be a field and $f$ a polynomial of positive degree $n$ with coefficients in $F$ and leading coefficient $a$. Let $f'$ be the derivative of $f$, $D$ the discriminant of $f$ and $R(f,f')$ the resultant of $f$ and $f'$. The theorem in question states that

$$R(f,f') = aD.$$

This theorem is already false for quadratic, monic polynomials over fields of characteristic 0. For example, if $f$ is $x^2 + x + 1$, the quadratic formula gives that $D = -3$, but $R(f,f') = 3$.

The error in the proof of the theorem comes about as follows. Let the factorization of $f(=f(x))$ over the algebraic closure of $F$ be

$$f(x) = a(x - r_1)\cdots(x - r_n).$$

Then one proves (the details appear in [3]), correctly, that

$$R(f,f') = a^{2n-1}\prod_{i\neq k}(r_i - r_k).$$

The error is caused by claiming that

$$\prod_{i\neq k}(r_i - r_k) = \prod_{i<k}(r_i - r_k)^2,$$

while the correct claim is of course that

$$\prod_{i\neq k}(r_i - r_k) = (-1)^{(n-1)n/2}\prod_{i<k}(r_i - r_k)^2.$$

Indeed, the correct theorem is

Theorem.

$$R(f,f') = (-1)^{(n-1)n/2}aD.$$

### References

1. N. Jacobson, Basic Algebra I, Freeman, 1974, p. 310, Exercise 1. Only monic polynomials are mentioned.
2. S. Lang, Algebra, Addison-Wesley, 1965, p. 139.
3. B. L. van der Waerden, Modern Algebra, vol. I, Frederick Ungar, 1949, p. 87.

---

## 139. MISCELLANEA

Experience has shown that the power of mathematical models in the study of the real world has increased along with the complexity of the models and the level of abstraction of the mathematics.

—N. Teodorescu (1968), quoted by K. Manteuffel in Wiss. Z. Tech. Hochsch. Magdeburg 23 (1979), 523–525.

# THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## TEACHING PROOFS: A LESSON FROM SOFTWARE ENGINEERING

MEL FRISKE

*Department of Mathematics and Computer Science, Wisconsin Lutheran College, Milwaukee, WI 53226*

**Introduction.** How can we provide a positive early experience with proofs for the average mathematical sciences student? The companion problem in computer science is how can the average computer science student design and write a computer program that works?

During the infancy of computing programmer productivity was unacceptably low, and the cost of modifying and maintaining poorly written software soared. In response the computing community developed software methodology [**3**], [**5**], [**6**] and brought it into the curriculum [**1**], [**2**], [**14**] with courses that emphasized technique as well as content (e.g., data structures, systems analysis and design, and software engineering). The results have been quite satisfactory, for students have responded well to the methodology.

There is a close analogy between the thought processes used in computer programming and those required for writing proofs. The methodology and principles which have been so effective in teaching programming can also be applied to teaching proof skills. Teaching proofs is one activity where mathematicians have had an unacceptable rate of success.

Traditionally a student's first intensive experience with proofs has been relegated to upper level courses in analysis and algebra. In such courses the primary focus is on the mathematical content. Consequently, proof related skills are often left for the student to absorb through some undefined process. The problem a mathematics student faces in proving a theorem is not much different from that which a computing student faces with a complex programming task. However, this intimidating barrier can be overcome by a systematic approach to problem solving. In computing there are three basic problem solving steps which we now consider.

**1. Problem Specification.** The first step is the hardest. How often have we not heard a mathematics student complain: "I just don't know where to start"? Computing students are taught that the first step in problem solving is problem specification: "Just what is the real problem?" One device used here is a data flow diagram [**12**] which diagrammatically represents the sources, flows, and destinations of all required data for the programming task. In mathematics, students usually learn the counterpart, namely, drawing a picture of the problem. Topologists are particularly adept at this.

Another computing aid is the IPO chart [**5**] which the problem solver uses to identify the desired output (O), the available (I), and the processing (P) steps required to transform the input into the output. Here there is a strong and valuable analogy to proof writing. To prove a theorem a mathematician first asks, "What do I have to show?" Then he asks, "What hypotheses and facts do I have to work with?" And finally, "How do I get from the given assumptions to the desired conclusion?" Notice the output, input, processing sequence once again. In teaching proof skills we need to identify and emphasize this three-step thought process.

For students the first of these questions is often difficult. How does one decide what has to be shown? In [**11**] this question is broken into three parts: the abstraction question, answering the abstraction question, and applying the answer to the specific problem. For example, suppose we are trying to prove that if $X$ is a probability space and $E$ is an event, then $P(E) \leqslant P(X)$. The abstraction question would be, "How can we show that the probability of an event is less than or equal to the probability of another event?" An answer to the abstraction question is, "Show that

the first event is a subset of the second event." Applying this to the problem at hand, we need to show that $E$ is a subset of $X$.

**2. Logical Design.** After problem specification a computing student knows that the next step is designing the program logic. This is normally done in abbreviated form without regard to details of language syntax. The result of this analysis is called pseudocode: a logically correct solution of the problem which, however, is in a form unacceptable for computer consumption. This logical design corresponds to a mathematician's initial sketch of a proof. The sketch is also abbreviated, contains the main steps, probably lacks some fine details, and requires conversion into acceptable English. We all know that this design step is the big challenge, this is where the blood is shed. Even here computing provides some methodological principles to ease the task.

The structured design principle is one developed by two mathematicians. In the early days of high level programming languages program style varied widely. Clever programmers would devise cute tricks, abbreviate freely, and construct complicated flows of control just to shave a second off of processing time [**13**]. The result was software which few people besides the author would ever want to read, software which often contained hard to find logical errors. What was needed was a small set of simple logical building blocks (control structures) from which any program could be built. In a remarkable theorem proved in 1964, Böhm and Jacopini [**4**] provided such a set. It contained only three structures: a sequence, a decision, and a loop. Although theoretically any computer program can be constructed from this small set, most programmers extend the set for practical purposes to include approximately seven control structures. A program designed exclusively from this set is then properly called a structured program.

Even freshmen can readily grasp this structured approach to the design of program logic. The same can be done with mathematical proofs. Certain basic logical techniques are used over and over in constructing proofs. In his book [**11**, pp. 97–98] Solow has identified what might be called a set of control structures for mathematical proofs. They include, for example, the "forward-backward" method, proving the contrapositive, proof by contradiction, construction proofs, induction, etc. Students can master the first few of these as sophomores, and as they mature, master others. It is not necessary to have a name for every technique, but it is important that students recognize a technique when they read it and be able to employ it themselves.

Two other design principles closely related to structured design are the top-down and modular approaches [**5**]. "Top-down" and "modular" refer to the manner in which a complex problem is broken into more manageable subproblems (modules). "Structured" refers to the set of logical building blocks from which the problem solution is constructed.

For example suppose that in graph theory we are asked to prove that graph isomorphism is an equivalence relation. After analyzing the problem our problem specification reveals that we need to show three things, namely, that graph isomorphism is reflexive, symmetric, and transitive. In a very natural way the big problem has been decomposed into three smaller subproblems (modules). For more examples and detail the reader should see the article by Leron [**8**].

**3. Writing the Program.** Upon completing the logical design a computing student converts it module by module into an appropriate programming language [**7**], paying special attention to the reserved words and syntax of the language. Converting a proof sketch into a paragraph proof is a similar task in which care must be taken with language and style.

If we truly wish to be more effective in teaching proof skills, we need to do a better job of teaching. There is no excuse: Pólya's classic work [**9**] has been with us for years, and his "How to Solve it" list beautifully elaborates the process of problem specification and the design principles mentioned earlier. To teach these ideas in computing we find it helpful to prepare a detailed syllabus indicating not only the material content but also a careful ordering of the methods, techniques, and principles to be taught and the types of programs the student will write. In mathematics we need to do the same. This includes specifying in advance the types of proofs students will be required to write and providing consistent, constructive criticism when grading them.

We also need new introductory texts providing systematic coverage of proof skills (for example [**10**]). Ideally this material can be interwoven with the mathematical content. Such texts might contain exercises requiring problem specification, a logical design, or analysis of a proof by breaking it into modules. Debugging exercises (What's wrong with this?) are also very effective.

SUMMARY. Software engineering offers a methodology and some principles for proof writing, namely,

A. Problem specification,
B. Logical design (structured, top-down, modular),
C. Writing the proof (attention to organization and special language).

A first significant exposure to these techniques should be provided in a lower level course which has some flexibility in mathematical content, so that sufficient time can be allocated to technique and not just content.

### References

**1.** D. Adams and T. Athey, eds., DPMA Model Curriculum for Undergraduate Computer Information Systems Education, DPMA Education Foundation Committee on Curriculum Development, Data Processing Association Education Foundation, 1981.

**2.** R. Austing, et al., eds., Curriculum '78: Recommendations for the undergraduate program in Computer Science—a report of the ACM Curriculum Committee on Computer Science, Comm. ACM, 22 (March 1979) 147–165.

**3.** F. Baker, System Quality through Structured Programming, in AFIPS Proceedings of the 1972 Fall Joint Computer Conference, AFIPS Press, 1972, pp. 339–344.

**4.** C. Böhm and G. Jacopini, Flow diagrams, Turing Machines, and languages with only two formation rules, Comm.. ACM, 9 (May 1966) 366–371.

**5.** B. Boehm, Software engineering, IEEE Trans. on Computers C-25 (December 1976) 1226–1241.

**6.** E. Dijkstra, Programming Considered as a Human Activity, in Proceedings of the 1965 IFIP Congress, North-Holland, 1965, pp. 213–217.

**7.** B. Kernighan and P. Plauger, Programming style: examples and counterexamples, ACM Computing Surveys, 6 (December 1974) 303–319.

**8.** U. Leron, Structuring mathematical proofs, this MONTHLY, 90 (1983) 174–175.

**9.** G. Pólya, How to Solve it, Princeton Univ. Press, 1945.

**10.** D. Smith, M. Eggen, and R. St. Andre, A Transition to Advanced Mathematics, Brooks/Cole, 1983.

**11.** D. Solow, How to Read and Do Proofs, Wiley, 1982.

**12.** W. Stevens, G. Meyers, and L. Constantine, Structured design, IBM Systems Journal, 13 (May 1974) 115–139.

**13.** N. Wirth, On the composition of well-structured programs, ACM Computing Surveys, 6 (December 1974) 247–259.

**14.** E. Yourdon and L. Constantine, Structured Design, Prentice-Hall, 1979.

## USING THE MULTIVARIABLE CHAIN RULE

FRED HALPERN
*10602 Stone Canyon Rd., #260, Dallas, TX 75230*

Finding the derivative of an expression with many occurrences of $x$ is a major application of the multivariable chain rule. The usual computation, using extraneous variables, introduces various side calculations and bookkeeping chores. This paper restates the multivariable chain rule to provide a simple straightforward method of computation. Informally, the computation states that

*The derivative of an expression with many occurrences of $x$ is the sum of the derivatives of the expression with respect to each separate occurrence of $x$.*

Thomas and Finney [**1**] (pp. 600–605) contains a standard statement of the multivariable chain

rule and its application to computing derivatives.

We consider functions $f(x_1,\dots,x_n)$ and are interested in $Df(x,\dots,x)$ where $D$ denotes differentiation with respect to $x$. Since $\partial x_i/\partial x = 1$ $(i = 1,\dots,n)$,

$$Df(x,\dots,x) = \sum \frac{\partial f}{\partial x_i}\bigg|_{(x,\dots,x)}.$$

When computing each $\partial f/\partial x_i$ the other variables are kept constant. This is signified by replacing the other variables by $c$, $x$ being resubstituted for $c$ after the differentiation.

With this agreement about the relation between $x$ and $c$, we obtain the following application of the multivariable chain rule to functions of the form $f(x,\dots,x)$.

THEOREM. *If each of*

$$Df(x,c,\dots,c),\, Df(c,x,\dots,c),\dots,Df(c,c,\dots,x)$$

*exists and is continuous in an interval containing* $c$, *then*

$$Df(x,x,\dots,x) = Df(x,c,\dots,c) + Df(c,x,\dots,c) + \cdots + Df(c,c,\dots,x).$$

The usual product and quotient rules for derivatives easily follow from the Theorem. In addition, an "exponential rule" for $Df(x)^{g(x)}$ can be obtained. Here are some applications of the Theorem.

EXAMPLES:

(1)
$$\begin{aligned} Dx\,e^x/\sin x &= Dx\,e^c/\sin c + Dc\,e^x/\sin c + Dc\,e^c/\sin x\\ &= e^c/\sin c + ce^x/\sin c + ce^c(-1/\sin^2 x)\cos x\\ &= e^x/\sin x + xe^x/\sin x - xe^x\cos x/\sin^2 x.\end{aligned}$$

(2)
$$\begin{aligned} Dx^{x^x} &= Dx^{c^c} + Dc^{x^c} + Dc^{c^x}\\ &= c^c x^{(c^c-1)} + (\ln c)\,c^{x^c} c x^{c-1} + (\ln c)\,c^{c^x}(\ln c\, c^x)\\ &= x^x x^{(x^x-1)} + x^x x^{x^x}\ln x + x^x x^{x^x}\ln^2 x.\end{aligned}$$

(3)
$$\begin{aligned} D\int_a^{f(x)} g(x,t)\,dt &= D\int_a^{f(x)} g(c,t)\,dt + D\int_a^{f(c)} g(x,t)\,dt\\ &= g(c,f(x))f'(x) + \int_a^{f(c)} \frac{\partial g}{\partial x}(x,t)\,dt\\ &= g(x,f(x))f'(x) + \int_a^{f(x)} \frac{\partial g}{\partial x}(x,t)\,dt.\end{aligned}$$

(4)
$$\begin{aligned} D\int_0^x\int_0^x f(u,v)\,du\,dv &= D\int_0^c dv\int_0^x f(u,v)\,du + D\int_0^x dv\int_0^c f(u,v)\,du\\ &= \int_0^c f(x,v)\,dv + \int_0^c f(u,x)\,du\\ &= \int_0^x f(x,v)\,dv + \int_0^x f(u,x)\,du.\end{aligned}$$

The Theorem also applies to expressions involving determinants, inner products, and cross products.

**Reference**

1. G. B. Thomas and R. L. Finney, Calculus and Analytic Geometry, 5th ed., Addison-Wesley, Reading, MA, 1980.

## ON REGULAR MARKOV CHAINS

NICHOLAS J. ROSE
*Department of Mathematics, North Carolina State University, Raleigh, NC 27650-8205*

The following basic theorem concerning regular Markov Chains is often presented in courses in finite mathematics:

THEOREM. *Let $P$ be the transition matrix for a regular Markov chain, then (i) $P^n \to W$ as $n \to \infty$, (ii) each row of $W$ is the same probability vector $w$ which has positive components, (iii) if $p$ is a probability vector, $p^{P^n} \to w$, (iv) $w = wP$.*

Since, generally speaking, students in such courses know little mathematics beyond high school algebra, the proof of the theorem is usually omitted. In this note we present an elementary proof in the case when $P$ is a $2 \times 2$ matrix.

Recall that a probability vector is a row vector with nonnegative components that sum to 1, a transition matrix is a square matrix whose rows are probability vectors, and a transition matrix for a regular Markov chain has the property that some power of the matrix has all positive entries.

A regular $2 \times 2$ transition matrix can be written in the form

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix},$$

where $0 < a \leqslant 1, 0 < b \leqslant 1, a + b < 2$. Before computing powers of $P$, we write $P = I + S$, where

$$S = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix}.$$

Letting $r = -(a+b)$, we find $S^2 = rS$, and $S^n = r^{n-1}S$ for $n \geqslant 2$. Computing the powers of $P$ we find

$$P^2 = (I+S)^2 = I + 2S + S^2 = I + (1 + (1+r))S,$$

$$P^3 = I + \left((1 + (1+r) + (1+r)^2\right)S,$$

$$\vdots$$

$$P^n = I + \left(1 + (1+r) + \cdots + (1+r)^{n-1}\right)S = I + \frac{1-(1+r)^n}{1-(1+r)}S.$$

Since $-1 < 1 + r < 1$, we find

$$P^n \to W = I + \frac{1}{a+b}S = \begin{bmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{bmatrix}.$$

This proves parts (i) an (ii) of the theorem; the other parts follow easily.

Finally we note that another way of computing $P^n$ is to use the binomial theorem:

$$P^n = (I+S)^n = \sum_{i=0}^{n} \binom{n}{i} S^i = I + \sum_{i=1}^{n} \binom{n}{i} r^{i-1} S$$

$$= I + \frac{1}{r}\left(\sum_{i=1}^{n} \binom{n}{i} r^i\right) S$$

$$= I + \frac{1}{r}\left((1+r)^n - 1\right) S.$$

# PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, DAVID BORWEIN (ADVANCED PROBLEMS), H. M. W. EDGAR (ELEMENTARY PROBLEMS), AND D. H. MUGLER



*Send all* ***proposed*** *problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions*** *should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3073. *Proposed by Calin P. Popescu, Bucharest, Rumania.*

In the hexagon $A_1A_2A_3A_4A_5A_6$ the triangles $A_1A_3A_5$ and $A_2A_4A_6$ are equilateral. Is it true that $A_kA_{k+3}$ $(k = 1,2,3)$ are congruent if their sum equals the perimeter of the hexagon? Also, consider the converse.

E 3074. *Proposed by M. J. Pelling, London, England.*

Prove the inequality

$$\sum(1+p^2)^2(1+q^2)^2(p-r)^2(q-r)^2 \geq 2\prod(1+p^2)\prod(p-q)^2, \qquad p,q,r \in \mathbb{R},$$

where the sums and products are over cyclic permutations of $p, q, r$, and determine all the cases where equality holds.

E 3075. *Proposed by Duane M. Broline, University of Evansville.*

Let $B_k$ be the $k$th Bell number, the number of partitions of $k$ objects. For $k \leqslant n$, show that

$$\sum_{\substack{a_1,a_2,\ldots,a_n\geqslant 0\\ a_1+2a_2+\cdots+na_n=n}} \frac{a_1^k}{1^{a_1}(a_1!)2^{a_2}(a_2!)\cdots n^{a_n}(a_n!)} = \sum_{\substack{b_1,b_2,\ldots,b_k\geqslant 0\\ b_1+2b_2+\cdots+kb_k=k}} \frac{k!}{(1!)^{b_1}(b_1!)(2!)^{b_2}(b_2!)\cdots(k!)^{b_k}(b_k!)} = B_k.$$

E 3076. *Proposed by F. S. Cater, Portland State University.*

Let $g$ be a continuous function from $R$ to $R$, let $f$ be a positive-valued continuously

differentiable function from $R^2$ to $R$, and let $F(x) = \int_0^{g(x)} f(x, y)\,dy$. We know that if $g$ is constant, we can differentiate under the integral sign to obtain the derivative of $F$:

$$(*) \qquad F'(x) = \frac{d}{dx}\int_0^{g(x)} f(x, y)\,dy = \int_0^{g(x)} D_1' f(x, y)\,dy \qquad (\text{all } x \in R).$$

Is the converse true in general? That is, if equation $(*)$ holds for all $x \in R$, must $g$ be constant?

E 3077. *Proposed by Nicholas Passell, University of Wisconsin–Eau Claire.*

Let $N$ be the natural numbers with the topology $T$ consisting of $\varnothing$, $N$ and the complements of finite sets. Characterize the continuous functions from $\mathbb{R}$, the reals with the usual topology, to $N$ with the topology $T$.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by June 30, 1985. The solver's full post-office address should be on each sheet.*

6488. *Proposed by Robert M. Young, Oberlin College.*

Show that if $\lambda_1, \lambda_2, \lambda_3, \ldots$ are the positive solutions of the equation $\tan x = x$, then

$$\sum_1^\infty 1/\lambda_n^2 = \frac{1}{10}.$$

6489. *Proposed by Eliot Jacobson, Ohio University.*

Suppose that $E$ and $F$ are fields such that the additive groups $(E, +), (F, +)$ are isomorphic, and the multiplicative groups $(E^*, \cdot), (F^*, \cdot)$ are isomorphic. Are $E$ and $F$ isomorphic as fields?

6490. *Proposed by J. Sutherland Frame, Michigan State University.*

Show that

$$\sum_{k=1}^\infty (k^2 + k)^{-3} = 10 - \pi^2,$$

and more generally that

$$(-1)^{n-1}\sum_{k=1}^\infty (k^2 + k)^{-n} = \binom{2n-1}{n-1} + \sum_{j=1}^{[n/2]} \binom{2n-1-2j}{n-1} B_{2j}(-4\pi^2)^j/(2j)!$$

where $B_2, B_4, B_6, \ldots$ are the Bernoulli numbers $1/6, -1/30, 1/42, -1/30, 5/66, \ldots$.

## SOLUTIONS OF ADVANCED PROBLEMS

### An Extension of Hadamard's Inequality

6430 [1983, 402]. *Proposed by J. Borwein, Dalhousie University, Halifax, Canada.*

Recall that the $p$-trace average of a positive definite symmetric $n \times n$ matrix $C$ (we write $C > 0$) is given by

$$j_p(C) := \left(\frac{1}{n}\sum_{i=1}^n \lambda_i^p\right)^{1/p}$$

for $-\infty < p < \infty, p \neq 0$. Here $\lambda_1 < \lambda_2 \leqslant \cdots \leqslant \lambda_n$ are the eigenvalues of $C$. Extend this

definition by continuity for $p = 0, \infty$, and $-\infty$. Thus

$$j_0(C) := (\det C)^{1/n}, j_\infty(C) := \lambda_n, \quad \text{and} \quad j_{-\infty}(C) := \lambda_1.$$

Let $\Delta(C)$ be the diagonal matrix whose diagonal coincides with the diagonal of $C$. Show that

(i) $$j_p(C) \leqslant j_p(\Delta(C)) \qquad \text{for } -\infty \leqslant p \leqslant 1,$$

and

(ii) $$j_p(C) \geqslant j_p(\Delta(C)) \qquad \text{for } 1 \leqslant p \leqslant \infty.$$

Note that for $p = 0$ this is the well-known Hadamard inequality.

*Solution by the proposer.* Let $\{D_k | k = 1,2,\dots,2^n\}$ be the set of diagonal matrices with entries of $\pm 1$. Then

$$\Delta(C) = 2^{-n} \sum_{k=1}^{2^n} D_k C D_k.$$

(See A. W. Marshall and I. Olkin, *A convexity proof of Hadamard's inequality*, this MONTHLY, 89 (1982) 687–688.) Moreover, $C$ and $D_k C D_k$ share the same spectrum. Since $j_p$ is concave (convex) for $p \leqslant 1$ ($p \geqslant 1$) we have

$$\begin{aligned} j_p(\Delta(C)) &\geqslant (\leqslant) 2^{-n} \sum_{k=1}^{n} j_p(D_k C D_k) \\ &= 2^{-n} \sum_{k=1}^{n} j_p(C) = j_p(C), \end{aligned}$$

and this establishes (i) and (ii).

Also solved by Michael Golomb; Olaf Krafft (West Germany); Marvin Marcus & John Bruno; Pei Yuan Wu (Republic of China); Richard P. Savage, Jr.; Alain Tissier (France); and Henry Wolkowicz (Canada).

### Irreducibility of the Poisson Distribution

6431 [1983, 402]. *Proposed by Gabor J. Szekely and Andras Zempleni, Budapest, Hungary.*

Show that if $X$ and $Y$ are independent nonnegative integer-valued random variables and $XY$ has a Poisson distribution, then either $X$ or $Y$ takes at most two values, zero and one, with probability one. (In this sense the Poisson distribution is irreducible.)

*Solution by L. E. Mattics, University of South Alabama, Mobile.* Let $m$ and $n$ be positive integers with $n > 1$. To be more general we replace the hypothesis that $XY$ has a Poisson distribution by the hypotheses

$$P(XY = m) > 0 \text{ for all } m,$$

and

$$\lim_{m \to \infty} P(XY = nm)/P(XY = m) = 0.$$

Since $P(XY = 1) = P(X = 1)P(Y = 1)$, both $P(X = 1)$ and $P(Y = 1)$ are nonzero. Let $q$ be a prime, then

$$P(XY = q) = P(X = 1)P(Y = q) + P(Y = 1)P(X = q).$$

Hence we can assume that (say)

$$P(X = q) \geqslant P(XY = q)/2P(Y = 1)$$

for infinitely many primes $q$. Consequently,

$$P(XY = nq) \geqslant P(Y = n)\,P(X = q)$$
$$\geqslant P(Y = n)\,P(XY = q)/2P(Y = 1)$$

for infinitely many primes $q$, and under the (new) hypotheses this implies that $P(Y = n) = 0$.

Also solved by Lajos Takàcs and the proposers.

### A $C^\infty$ Function Supported on a Given Open Set

6438 [1983, 486]. *Proposed by J. L. Brenner, Palo Alto, California.*

Construct a differentiable real function that is positive at every rational point, but zero at an uncountable number of points, or prove that no such function exists.

*Solution by Ivan Netuka and Jiří Veselý, Prague, Czechoslovakia.* The following construction due to L. E. May (*Canad. Math. Bull.*, 12 (1969) 25–30) shows more generally that given any open set $G$, there is a $C^\infty$ function that is positive on $G$ and zero elsewhere. The set $G$ is the union of disjoint intervals $(a_n, b_n)$, $n = 1, 2, \ldots$ . Define

$$f_n(x) = \exp\left(-(x - a_n)^{-2}(b_n - x)^{-2}\right)$$

for $x \in (a_n, b_n)$, and $f_n(x) = 0$ elsewhere. Let

$$M_n = \sup\left\{\,|f_n^{(k)}(x)| : x \in \mathbb{R}, 0 \leqslant k \leqslant n\right\}.$$

Then $f = \sum_{n=1}^{\infty}(f_n/n^2 M_n)$ is positive on $G$ and is zero elsewhere, and it is easily seen that $f \in C^\infty$.

Also solved by 28 others.

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Books IV to VII of Diophantus' Arithmetica: In the Arabic Translation Attributed to Qusṭâ Ibn Lûqâ.* By Jacques Sesiano. Sources in the History of Mathematics and Physical Sciences Number 3. Springer-Verlag, New York, 1982. xii + 502 pp.

WILBUR KNORR
*Department of Classics/Philosophy Program in the History of Science, Stanford University, Stanford, CA 94305*

If Diophantus had not existed, no historian of ancient Greek mathematics would have invented him. Unlike the formal geometric treatises of Euclid, Archimedes and Apollonius, taken as representative of the characteristically Greek style of mathematics, Diophantus' *Arithmetica* falls within the category of arithmetic theory. Yet as antiquity's foremost crypto-algebraist, Diophantus is disappointing for his failure to treat this subject as one would later come to do. His aim is to present the solutions to arithmetic problems, via methods of varying degrees of generality, not the systematic exposition of a subject matter. Although he progresses from simpler to more complex problems, he seems not to attempt to define any clear lines of deductive structure in his work. One might wish to describe it as investigating the solutions to determinate and indeterminate systems of equations of first, second and third orders, but that conforms only tenuously with Diophantus' selection and ordering of problems (cf. the survey list compiled by T. L. Heath [*Diophantus*, 1910, 260–266]). Moreover, in contrast with the modern field of "Diophantine" analysis, Diophantus

allows rational solutions, not just integral solutions, thus drastically reducing the difficulty, and hence the interest, of many of his problems. He is thus never called on to develop those methods which grace the finest of the Indian efforts in indeterminate analysis; indeed, the indeterminate first-order problem in two unknowns is so trivialized as not to merit inclusion at all.

The fact that only six of the original thirteen books of the *Arithmetica* survive in Greek manuscripts had led scholars to speculate that the lost portions contained the more interesting and advanced materials. With the appearance of Jacques Sesiano's edition of the Arabic version of Diophantus' *Arithmetica*, we should have expected enlightenment on this issue. Instead, we encounter new puzzles, these primarily of the textual sort. First, the Arabic version presents itself through its own titles as being Books IV through VII of the *Arithmetica*; but comparison with Books IV through VI in the Greek reveals at once that no Book in the one sequence duplicates any in the other. (In view of this, I shall refer to the Arabic Books as IVA, VA, etc., and the Greek Books as IVG, VG, VIG.) Second, there are important stylistic differences, the Arabic version tending toward prolixity quite out of keeping with the austerely terse wording in the Greek.

Either of these considerations could raise doubts of the authenticity of the Arabic Books. Sesiano establishes convincingly that the Arabic does indeed transmit a portion of Diophantus' work, specifically, the original Books IV through VII [Sesiano, pp. 4–8]. For instance, Book IVA bears a preface stating that its project is to extend to the solid numbers the methods that the previous Books had applied to linear and planar numbers; thus, rather than introduce new methods, the following Books propose to inculcate skill in the manipulation of techniques already presented. As Sesiano notes, this is a fair characterization of the Arabic Books, for they adhere to the technical base established in Books I-III. By contrast, Book IVG begins at once with cubic problems handled via new methods. Clearly, Books IVA-VIIA would be poorly suited for placement after IVG. In this way, speculations among older scholars about the scope of the missing portions of Diophantus' work are partly settled [Sesiano, sect. 13; cf. Heath, pp. 6–12]: Tannery's hope to discover interesting advanced materials in the lost Books of the *Arithmetica* are effectively dashed, as we realize that the four recovered Books represent no increment in technical level. On the other hand, the conjecture of Nesselmann and others, that the lost Books fit somehow within the sequence of those extant in Greek, has in part been sustained; but the gap turns out to be between Books III and IVG, rather than between Books II and III as they had surmised.

On a second question, relating to the stylistic differences between the Arabic and Greek Books, Sesiano's proposed explanation is, in my view, not altogether persuasive. Exploiting the ancient testimony that the mathematician Hypatia (d. 415 A.D.) had composed a commentary on Diophantus, Sesiano suggests that the Arabic Books were translated from a Greek text which had incorporated substantial materials drawn from some such "major commentary," whereas, by contrast, the extant Greek recension represents a text much closer to the one produced by Diophantus himself [Sesiano, sect. 9, 10, and 12.2]. This view is intended to account for systematic differences of style which distinguish the Arabic version from analogous parts of the Greek. Even if the Arabic translators introduced certain changes of their own, as Sesiano allows, he maintains that the more important differences must be referred to their Greek source.

Among the minor sort of discrepancies due to the translator is the manner of rendering numerical terms; the Greek employs a notational system comparable to our own [**1**], whereas the Arabic writes out in full the number-words; thus, the Arabic of VIA, 4 expresses the number 2563001 as "two thousand thousands and five hundred and three and sixty thousand and one," where the Greek must have held a numeral like "256 My 3001" ($\sigma\nu s M^{\upsilon}\gamma\alpha$, where "My" stands for *myriades*, or "ten-thousands"). So too, the translator must be responsible for replacing Diophantus' "syncopated" mode with the full "rhetorical" rendition of algebraic expressions. Thus, in VA,7 we read, "The sum of these two cubes is two thousand units and sixty powers [sc. $2000 + 60\ x^2$],... equal to two thousand and two hundred and forty units;" a comparable passage in IVG,1 reads, "Then the cubes will be Dy 30 Mo 250 [sc. $30\ x^2 + 250$]; these are equal

to Mo 370" [2]. In the standard view of the progression from "rhetorical" through "syncopated" to "symbolic" expressions in algebra, this might seem to be a backward step [3]. But in this the translator has correctly intuited that the lines in the Greek were intended to be read out in full, even if the text is highly abbreviated. His practice might thus be viewed as an aid to the student.

Sesiano admits that these differences in style are due to the translator; but he wishes to view other aspects of style as borrowings from the "major commentator" built into the text used by the translator. Among these are the belabored algebraic manipulations in the Arabic, where comparable passages in the Greek omit intermediate steps; the insertion of full statements of algebraic identities in the Arabic, where they would be tacitly assumed in the Greek; and the full working out of the "proof" in the Arabic, where the Greek merely states the solutions. Sesiano argues that since amplifications of these sorts are encountered in other Arabic renderings of Diophantus, in particular, those of al-Karajî, apparently independent of the present translation by Qusṭâ, these must have been held in the Greek prototype.

To support his view, Sesiano calls attention to several particular instances where comparable features in the Arabic indicate a basis in the Greek [Sesiano, sect. 10]. For instance, in VIA,4 one of the numbers of the solution is stated as 2563001 (that is, written out in full, as noted above); but the correct answer is in fact the product of 251 times 63001. The slip seems possible only through the mediation of a numeral notation like that employed in the Greek of Diophantus. Thus, an expression like "251 times 6 My 3001" might be wrongly construed as "256 My 3001" via a scribal deletion, either by a Greek scribe or the Arabic translator. What seems clear, however, is that the Greek text must at one point have included the expression "251 times 63001," and this is precisely the kind of additional detail in the working out of the solution which distinguishes the Arabic from parallels in the Greek.

A first reservation one may have about this use of the hypothesis of the "major commentary," however, is that the sorts of changes at issue here are of such a low level as not to require any real mathematical insight. Retrieving the commentator via stylistic comparisons of the Arabic and Greek texts would thus isolate an essentially trivial mind. This is in direct conflict with ancient testimonies of Hypatia's high caliber as a philosopher and mathematician. Given that we possess not a single directly attested word from her works, I think that historical fairness, if nothing else, should lead us to resist this project of reconstruction.

But perhaps this is mere sentimentality. A consideration of certain textual issues, however, occasions deeper misgivings. Sesiano would maintain that the manuscripts available to the 9th-century Arabic translators held a revised edition of Diophantus, influenced by the commentary of Hypatia, whereas the manuscripts available to the 13th-century Byzantine scholars, principally Maximus Planudes, on which the extant Greek tradition is based, represented the older, pre-Hypatian form of the text [4]. But, as a rule, new recensions of technical works largely displaced the older versions. The textual history of works of Euclid and Apollonius, for instance, followed this basic pattern [5]. Sesiano's position thus conflicts with our evidence relating to the transmission of ancient technical works.

I cannot hope to develop here an alternative view of the relation of the Greek and Arabic versions of Diophantus, but perhaps I can indicate the direction one might pursue. The Planudean manuscripts present not only a text of Diophantus, but also extensive scholia, attributed to Planudes himself, amounting to a running commentary on Books I and II. These scholia appear sometimes at the ends of the propositions they relate to, sometimes in the margins, and attempt to amplify the text, for instance, by filling in the intermediate steps of algebraic resolutions, inserting the statements of identities assumed in the text, working out in full the proofs of the solutions, and so on. These are comparable to the kinds of amplification found in the Arabic version of Diophantus. Although elaborations like these are obvious enough, we are no more compelled to suppose that Planudes originated them all in his commentary, as to suppose that Qusṭâ originated the same in his translation. We can as readily conceive that the manuscripts at Planudes' disposal already contained interpolated comments of this sort, and that he gleaned from these the materials of his commentary, occasionally adding remarks of his own. We would suppose, further, that the

earlier manuscripts were in comparable condition, so that the 9th-century Arabic translators had before them both the text and its accompanying marginal commentary. They would of course recognize the pedagogical value of this marginal material, and so, quite naturally, could attempt to incorporate it into their translation. In this way, the Arabic and Greek recensions would owe their stylistic differences not to essential differences in the source texts utilized, but rather to differences in editorial technique.

Although I have thus expressed doubts about one of Sesiano's central interpretive positions, his views over all appear to be quite sound. More important for the practical utility of his edition is his care to separate his interpretations from the body of the text and translation. In this he parts company with the major 19th-century editors, whose penchant for bracketing purported interpolations, for instance, now enormously complicates the use of their editions. The reader can thus consult Sesiano's text without concern over prejudgment, yet if he choose, can readily locate the places in the introduction and commentary where Sesiano presents his own views. The translation is conveniently cross-referenced to the text via skillful use of line numbers. The introduction presents, among other things, a nice résumé of the textual history of arithmetic theory in Greek and Arabic. Useful study aids include a list in modern notation of problems from the Greek and Arabic Books of Diophantus and an Arabic lexicon, with frequent noting of Greek equivalents. Doubtless, many readers would have preferred a more extensive noting of English equivalents than appears in this lexicon, and this is, regrettably, but one of several ways in which Sesiano fails to make accommodations for the general reader. His edition will thus not readily serve as an introduction for the nonspecialist. Nevertheless, used in conjunction with the Greek edition of Tannery and the standard translation-commentaries (e.g., those by Heath, Ver Eecke and Czwalina), Sesiano's work will provide a valuable instrument for renewed Diophantus studies.

With reference to Sesiano's ample bibliography, two salient omissions should be noted. First, F. Sezgin's *Geschichte des arabischen Schrifttums*, V: *Mathematik* (Leiden, 1974) has become an indispensable reference resource for the study of Islamic mathematics, and provides capsule accounts and bibliographical guidance to manuscripts, editions and secondary literature for most of the mathematicians mentioned by Sesiano. Second, there are no citations of contributions by Roshdi Rashed, save for his edition of the algebra of Samaw'al (co-edited with S. Ahmad, 1972). Among the pertinent items would be Rashed's article on al-Karajî (*Dictionary of Scientific Biography*, 7, pp. 240–246); his survey of the Arabic recension of Diophantus (*Revue d'histoire des sciences*, 1974, 27:97–122 and 1975, 28:3–30); and his Arabic edition with Arabic commentary of the same manuscript now edited by Sesiano (Cairo, 1977). It is clear from Sesiano's review of Rashed's edition (*Isis*, 1977, 68:627–630) that he finds much therein to criticize. But Sesiano's present silence on these prior efforts is surely a poorer solution to a delicate critical problem than the one he adopted in his earlier essay on Diophantus (*DSB*, 15, pp. 118–122), and I wish he had reconsidered this extraordinarily ungracious gesture.

### Notes

1. The Greek numeral system employs the consecutive letters of the alphabet to stand for the digits from 1 to 9, the tens from 10 to 90, and the hundreds from 100 to 900; a numeral will thus consist of the figures from each order which sum to the desired value. Auxiliary strokes and signs permit the expression of thousands and ten-thousands. Although lacking the place notion, it is comparable to our own system for the purposes of rational operations. For an account, see T. L. Heath, *A History of Greek Mathematics*, Oxford, 1921, I, pp. 31–40.

2. The abbreviations parallel those adapted in the Greek: Mo = forms of *monas* ("unit"); Dy = *dynamis* ("power," that is "second-power," or "square"); the Greek usually expresses addition via simple juxtaposition.

3. This view of the evolution of algebra is notable in G. H. F. Nesselmann's *Die Algebra der Griechen*, Berlin, 1842 and is elaborated by J. Klein in his *Greek Mathematical Thought and the Origin of Algebra*, Cambridge, Mass., 1968.

4. Sesiano [pp. 16–20] and Heath, *Diophantus*, pp. 14–19.

5. Euclid's *Elements*, Apollonius' *Conics* and Ptolemy's *Syntaxis*, for instance, led to the eventual extinction of

their precursors. The editions of Euclid's *Elements* and *Optics* by Theon of Alexandria (4th century A.D.; father of Hypatia) overwhelmed the tradition of these works, without entirely supplanting the older recensions. The commentaries by Eutocius of Askalon (early 6th century) affected the tradition of Archimedes' and Apollonius' works. This phenomenon of displacement is neither surprising nor deplorable. Most of these writers were engaged in the preparation of textbooks serving their students' and colleagues' needs, not of philologically pure texts. By implanting these revised versions into their own tradition of teaching, they merely secured their survival at the expense of the earlier versions.

---

*Selecta*: *Expository Writing*. By P. R. Halmos. Springer-Verlag, New York, 1983. xix + 304 pp.

R. P. Boas
*Department of Mathematics, Northwestern University, Evanston, IL 60201*

There are often said to be analogies between mathematics and music, but it is less often realized that there are analogies between departments of mathematics and of music. Consider the complaint of the music critic Andrew Porter (*The New Yorker*, August 16, 1982) about a contemporary composer whose work appeals to the "middle-brows," people who "know what they like and like what they know." His worry is that if these Philistines get what they want, standards will fall and progress in music will eventually cease. The parallel with mathematics is rather clear, if we identify the Philistines with the users of mathematics, whether they are the students who do not intend to become creative mathematicians, or the people who want to use mathematics in solving their own practical problems.

As a musical Philistine, I perceive weaknesses in Porter's position. Most obviously, musicologists are paid to devote time to studying avant-garde music and learning to appreciate it. Unfortunately for the rest of us, they do not spend much time explaining to us what it is all about (they are not paid to do that). We are not helped by just being asked to listen to the stuff.

The mathematical Philistines may reasonably make the same complaint about the avant-garde mathematicians who are (quite understandably) more interested in advancing mathematics than in explaining it. Even those who understand it and could explain it are usually inhibited because the folks who give out money are interested only in published research, not in exposition. Halmos (in one of the essays in this book) makes the point that doing research is important for teachers—but he doesn't say (or even imply) that it should be publishable research. The president of my university, on the other hand, is on the record as saying that "publish or perish" means that if you don't publish research, you will automatically perish as a teacher. (This is clearly an over-simplification, if only because most professors at most universities have little chance to teach anything except a pretty standard curriculum.) It is no wonder that modern mathematics (like modern music) has the reputation of being abstruse, incomprehensible, and remote from reality. There are other barriers against the good exposition of mathematics, but readers of this journal can think of them for themselves.

It is at least plausible that students of mathematics, and especially the users of mathematics who apply new mathematics—even without having read the proofs—to solving practical problems, would benefit from more good exposition, but few creative mathematicians spend much time explaining what they and others have done. There are some honorable exceptions, of whom Halmos is an outstanding example. Not only is he unafraid of having people look down on him for explaining advanced mathematics in a comprehensible way, he evidently enjoys explaining it, and he does it superbly. In his book you will find not only samples of his expository writing at various levels, but instructions on how to do it, as well as inspirational sermons and stimulating (if possibly irritating) comments on the mathematical scene. I regret that he did not choose to include his spirited defense of complex analysis, written in response to allegations that the subject is outmoded (Notices Amer. Math. Soc., vol. 18 (1971) 69). If more mathematicians would follow

Halmos's example and attend to his precepts, there might be less mathematical writing, but it would be more comprehensible. There is—most of us know it but are seldom willing to admit it—such a thing as good mathematical style, and Halmos is one of the small number of its exemplars. Whether you are a serious mathematician who is inclined to look down on "middle-brows," or you are looking for information about some mathematics beyond what you were taught or what you specialize in, or you are willing to listen to a master stylist explain how to write, or you just want to be entertained by the conversation of an interesting personality—read this book!

---

*Group Analysis of Differential Equations*. By L. V. Ovsiannikov. Academic Press, Inc., New York, 1982. xvi + 416 pp. $54.00.

WILLARD MILLER, JR.
*School of Mathematics, University of Minnesota, Minneapolis, MN 55455*

Group analysis of differential equations was inaugurated and developed by the Norwegian mathematician Sophus Lie in the closing decades of the nineteenth century. It is based on one of the most useful concepts in mathematics, that of a group $G$ acting on a manifold $M$: a transformation group. (For $p \in M$, $g \in G$ we denote the group action by $p \to pg \in M$.) Associated with this concept are such intuitive notions as the orbit of a point $p \in M$ (the set of all points $pg$ in $M$ as $g$ runs over all group elements) and an invariant point $p$ ($pg = p$, $\forall g \in G$). Furthermore, a function $f$ on $M$ is invariant under $G$ if $f(pg) = f(p)$ for all $p \in M$, $g \in G$ and the equation $f(p) = 0$ is invariant under $G$ if $f(pg) = 0$ whenever $f(p) = 0$ for all $g \in G$, i.e., $pg$ is a solution whenever $p$ is a solution.

In order to apply these notions to differential equations, the basic concepts are modified and specialized in three ways. First it is assumed that $G$ is a real analytic group (a Lie group) and that the coordinate transformations $p \to p' = pg$ are $C^\infty$ functions. This permits the application of ideas from calculus and leads to the construction of Lie algebras of Lie derivatives (infinitesimal group transformations). Second, increased generality is obtained by merely assuming that $G$ is a neighborhood of the identity element in an analytic group and that $M$ is a local coordinate patch with real coordinates $\mathbf{z} = (z_1, \ldots, z_m)$ for $p$. These two assumptions amount to the statement that $G$ is a *local Lie transformation group* on $M$. Finally, it is assumed that $M = X \times Y$ where the local coordinates for $X$ are the *independent variables* and the local coordinates for $Y$ are the *dependent variables*. For example, if $\dim X = n$, $\dim Y = 1$ coordinates of point $p$ would be $(x_1, \ldots, x_n, y) = (\mathbf{x}, y)$. Now suppose $y = u(\mathbf{x})$ for some $C^\infty$ function $u$ and consider the group transformation $(\mathbf{x}', y') \equiv (\mathbf{x}, y)g$. It is a straightforward consequence of the chain rule that the action of $G$ as a transformation group can be extended to the manifold $M_k = X \times Y_k$, where the dependent coordinates in $Y_k$ can be taken as $(y, y_{j_1}, y_{j_1 j_2}, \ldots, y_{j_1 \cdots j_k})$, i.e., the partial derivatives of $y$ up to order $k$. This extension of the group action to $k$th order derivatives of $y$ is known as the $k$th order *prolongation* of $G$.

A $k$th order differential equation in $y$ can be considered as an equation $f(p_k) = 0$, $p_k \in M_k$. We say that $G$ is a *symmetry group* for this equation provided the equation is invariant, i.e., provided $p_k g$ is a solution for each $g \in G$ whenever $p_k$ is a solution. Knowledge of a symmetry group $G$ for a differential equation can prove extremely useful. Given any solution of the equation one can generate an orbit of additional solutions by application of the symmetry transformations. Special solutions can be obtained through the requirement that they are invariant or partially invariant under the action of $G$. (Many of the special functions of mathematical physics arise exactly in this manner and a knowledge of $G$ is crucial in understanding the properties of such functions.) Through the classical theorem of Emmy Noether (1918) and its modern generaliza-

tions, one can use certain symmetries of Euler-Lagrange equations to obtain conservation laws for these equations. (In particular this method can be used to generate an infinite number of conservation laws for the Korteweg-de Vries equation.) The technique of separation of variables for second order linear partial differential equations can be completely understood in terms of symmetries, as can the technique of dimensional analysis for the equations of mechanics. In short, Lie's symmetry approach to differential equations has turned out to be enormously useful. Yet the full ramifications of this approach are far from exhausted and new work appears apace.

Among the contributors to this field in the past 25 years no one has been more influential than Ovsiannikov. In order to apply group theoretical methods to obtain information about the solutions of a given differential equation, one needs to know the "maximal" symmetry group of the equation. Lie solved this problem through the use of infinitesimal generators for the group action and reduced the computation of the symmetries of a fixed (nonlinear) differential equation to the solution of a system of linear differential equations. Ovsiannikov's work has centered on the refinement of Lie's methods in order to determine the detailed structure of the maximal symmetry groups for partial differential equations of physical interest, and especially to classify the invariant and partially invariant solutions of these equations.

This book (a translation of the 1978 Russian edition) contains a complete exposition of the basic Lie theory of local transformation groups and the determination of the groups which can be the symmetry groups of differential equations. A novelty of the treatment is that Lie's three fundamental theorems on transformation groups are proved in the more general setting of Banach (generally infinite dimensional) Lie groups and algebras. The survey of applications is limited primarily to the construction of partially invariant solutions, classification of second-order linear equations, conservation laws and other areas where the author has made personal contributions.

This present book is a considerable extension and updating of the author's influential 1962 monograph "Group Properties of Differential Equations" (in Russian). A generation of English speaking applied mathematicians studied this monograph through the help of a 1967 translation by G. W. Blumen which was circulated in mimeographed form. The greater degree of generality and abstraction (particularly the use of Banach Lie groups) employed by the author make the updated work less accessible to beginners than the 1962 original. Nevertheless, the general availability of Ovsiannikov's work in English translation is a very welcome development for wide dissemination of this beautiful and useful theory.

---

*The Lady or the Tiger? and Other Logic Puzzles*. By Raymond Smullyan. Knopf, New York, 1982. ix + 226 pp.

STEWART SHAPIRO
*Department of Philosophy, Ohio State University, Newark, OH 43055*

As indicated by the title, this is a book of logical and mathematical puzzles. Its author, an eminent logician, is developing a much deserved reputation among the general public as a puzzle-master (thanks, in part, to Johnny Carson). I confess a previous addiction to logic puzzles and hereby admit that I thoroughly enjoyed the book.

In this review, I would like to address a rather strong assertion made in the Preface (p. vii):

> So many people I have met claim to hate math, and yet are enormously intrigued by any logic or math problem I give them, provided I present it in the form of a puzzle. I would not be at all surprised if good puzzle books prove to be one of the best cures for so-called "math anxiety." Moreover, any math treatise *could* be written in the format of a puzzle book!

I would tend to agree in part, and disagree in part, with this pedagogical statement. First, the

disagreement: I doubt that puzzles will help those with severe cases of math anxiety. My experience is apparently different from that of Professor Smullyan. Among people I know, virtually all who enjoy logic puzzles also enjoy mathematics (and are good at both). Those who hate math also hate puzzles. In fact, an attempt to introduce a logic puzzle to a friend who suffers from math anxiety typically produces confusion, frustration, and unsolicited statements like "You know I hate math." Often, he or she will even have trouble following the given solution. The reason for this, of course, lies in the close analogy between puzzles and mathematics and may shed some light on the nature and causes of math anxiety. In principle, both mathematical discourse and logic puzzles involve long, directed, and focused stretches of careful reasoning from clearly stated premises. In this way, both differ from ordinary nonmathematical thinking. Moreover, some logic puzzles (especially those of the present book) involve concepts nearly as abstract as those of mathematics.

The extent of my agreement with the above quoted passage is also due to the close relationship between mathematics and logic puzzles. Professor Smullyan's books themselves demonstrate that certain mathematical ideas can be presented (to those receptive) through puzzles in an entertaining and engaging manner. One of the purposes of the book presently under review is to illustrate Gödel's incompleteness theorem and Tarski's theorem on the formal undefinability of truth. I found this a fascinating, thought-provoking, and insightful account and admit that I learned from it, but it is hard to judge the extent to which an average lay reader can grasp the mathematical and philosophical conclusions of this aspect of the book. It is clear, however, that certain more basic mathematical concepts and techniques are given lucid exhibitions in this puzzle book. In several places, for example, many instances of a certain type of problem are presented and the solutions are carefully shown to be similar. Then a "meta-problem," which generalizes the previous work is formulated and solved. The reader thus encounters the mathematical technique of generalization, the notion of "theorem," and, through the solutions, the notion of rigorous deductive proof. I suggest, then, that puzzle books such as this can go a long way toward illuminating at least some of the essential elements of mathematical activity.

It might be added that such books can also have philosophical value. In particular, many of the sections of this book help shed light on the common notions of truth, knowledge, and belief by drawing out interesting and counterintuitive consequences of several prima facie plausible doxical and epistemic doctrines.

---

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

In their paper, "Newton's method, circle maps, and chaotic motion", published in the January issue of this MONTHLY, Donald Saari and John Urenko study the convergence properties of Newton's method applied to Real Polynomials with Only Real Roots (RPORR) and to polynomial-like functions. Their results are generalizations of Barna's theorem: *If $f$ is* a RPORR, *then the set of nonconverging points $S_3$ is a Cantor set with Lebesgue measure zero.*

In [**1**], we stated a theorem which gives more information on the dynamics of the Newton transform $N$ on $S_3$. The proof is based on a symbolic representation closely related to the representation introduced in Saari's and Urenko's paper.

Let $A$ be an alphabet having $n - 1$ symbols and let $S$ be the set of sequences formed using elements in $A$ such that two consecutive elements are different. We show that the restriction of $N$ to $S_3$ is topologically equivalent to the shift acting on $S$. This can be extended to Saari's and Urenko's results.

Two important facts can be deduced from all these results:

1. The Newton transform of two $n$th degree RPORR are topologically equivalent.
2. Newton's method applied to RPORR converges almost everywhere.

These two remarks lead to very efficient implementations of an algebraic equations solver based on Newton's method.

It is commonly thought in the numerical analysis of algebraic equations that the higher the order of convergence of a method, the narrower is its domain of convergence. In case of Newton's method applied to RPORR, the order is two and the domain is almost all of $R$. Contrary to the common perception, however, there exists a globally convergent third order method-Laguerre's method [**2**]:

$$X_{i+1} = L(X_i) = X_i = \frac{nP(X_i)}{P'(X_i) + \Sigma H(X_i)},$$

where $\Sigma = \text{sign } (P'(X))$ and $H(X) = (n-1)^2 P'^2(X) - n(n-1)P(X)P''(X)$. This method applied to RPORR converges for every starting point in $R$. We do not know if there exists a fourth order method globally convergent for such polynomial equations.


**References**

**1.** M. Cosnard and C. Masse, Convergence presque partout de la méthode de Newton, C. R. Acad. Sci., 297 (1983) 549–552.

**2.** M. Leuze, A hybrid Laguerre Method, BIT 23 (1983) 132–138.



M. Cosnard–C. Masse
Laboratoire TIM 3
Institut IMAG
Université de Grenoble—BP68
38402 St-Martin-d'Hères cedex
France


---

## 140. MISCELLANEA

### CLEAR THEORY


JEREMY GRAY
*The Open University, Milton Keynes MK7 6AA, England, United Kingdom*


Explanations (lectures, articles, or books) in mathematics come in a variety of forms. As a discrete attempt to impose a valuation upon them, one is naturally led to make the following distinctions.

An explanation is *globally clear* if the overall structure of the argument is apparent; it is *locally* clear if it can be followed line by line.

An argument that is locally clear except for a small (finite?) set of deductions may be said to be *generically* clear. An argument may be said to be *asymptotically* clear if it becomes clearer upon

repetition. Rather worse are the arguments that are stochastically clear; these are arguments that are only occasionally locally clear and then only clear apparently at random.

An argument is *uniformly* clear if it is clear for everyone. A stronger condition is that an argument be *stably* clear; this is the case when its clarity survives repetition by an original member of the audience. The clarity of an argument that does not survive indefinite repetition will be said to be *unstable*; if, however, something always survives, the clarity may be said to be *semi-stable*.

It is often the case that an argument only becomes clear on localizing at an example; one may then say the argument contains the *germ* of a true idea. Or, by severely abusing notation, one may say the argument has a *local ring* to it.

---

## 141. MISCELLANEA

### LISP: THEY MAY ALSO SERVE, WHO DECLINE TO ENLIST...

ROBERT M. BAER
*379 Countyview Drive, Mill Valley, CA 94941*

Last year we were told that "Lisp is crisp" [1], and this year we are told that LISP "is the only language in which civilized people write computer programs" [2]. (There is a counter-view to the effect that civilized people view with distaste any language in which assignment and equality are represented by the same symbol.) LISP has recursion, but its overhead is notorious. Its primary (perhaps only) virtue is that list-processing capability (but, then, there are all those parentheses...). In any case, we put the question to our desktop verse-processor, and got the following output:

**Adrift on seas of parentheses**

LISP isn't crisp but essentially soggy;
Parentheses piled up make you feel groggy.

Deciphering LISP would apparently be
More work than Minoan Linear B.

As for cond (aka if) and cdr and car,
Like heavy seas they're best viewed from afar.

And vertigo thrives in anticipation
Of waves upon waves of Polish notation.

LISP for list-processing: cause to abort.
(Is it listing to starboard or listing to port?)

Yes, recursion is nice
To put puzzles on ice.

(But they dole out machine time, ration by ration,
So I'll skip the recursion and use iteration.)

**References**

1. D. R. Hofstadter, Metamagical Themas, Scientific American, 248 (1983), #3, 22–30; #4, 14–28.
2. M. Wand, What is LISP?, this MONTHLY, 91 (1984) 32–42.

# *New in 1985*

## *DEVELOPMENTAL TEXTS*

***BEGINNING ALGEBRA** by R. David Gustafson and Peter D. Frisk, Rock Valley College.* *Offers a broad range of topics at a depth appropriate for students with no prior experience with algebra.*

***INTERMEDIATE ALGEBRA FOR COLLEGE STUDENTS, 3RD EDITION** by Karl J. Smith and Patrick Boyle, Santa Rosa Junior College.* *Motivates new techniques and concepts within the setting of concrete examples, often from everyday experience.*

***INTERMEDIATE ALGEBRA** by Gene Sellers, Sacramento City College.* *Presents new concepts and techniques in an informal, non-threatening way.*

***ALGEBRA FOR COLLEGE STUDENTS** by R. David Gustafson and Peter D. Frisk, Rock Valley College.* *Designed for courses that combine topics from intermediate algebra and college algebra and for those who want to use one book for both courses.*

## *PRECALCULUS*

***PLANE TRIGONOMETRY, 2ND EDITION** by R. David Gustafson and Peter D. Frisk, Rock Valley College.* *Highly applied and concretely motivated.*

***COLLEGE ALGEBRA, 3RD EDITION** by Karl J. Smith and Patrick J. Boyle, Santa Rosa Junior College.* *New and ideas unfamiliar to the student are motivated by familiar images.*

***ALGEBRA AND TRIGONOMETRY, 3RD EDITION** by Bernard Rice and Jerry Strange, University of Dayton.* *Emphasizes mastery of algebraic and computational skills with its traditional approach to trigonometry.*

***COLLEGE ALGEBRA AND TRIGONOMETRY** by Robert P. Webber, Longwood College.* *Focuses on doing problem solving to understanding mathematical relationships.*

## *COMPUTER MATHEMATICS*

***ELEMENTS OF COMPUTER MATHEMATICS** by Sandra Talbot, Mattatuck Community College; Harold Baker, Waterbury State Technical College; and Lawrence Gilligan, University of Cincinnati.* *A practical approach emphasizes basic concepts and skills.*

# The Mathematical Association of America

## CORPORATE MEMBERS

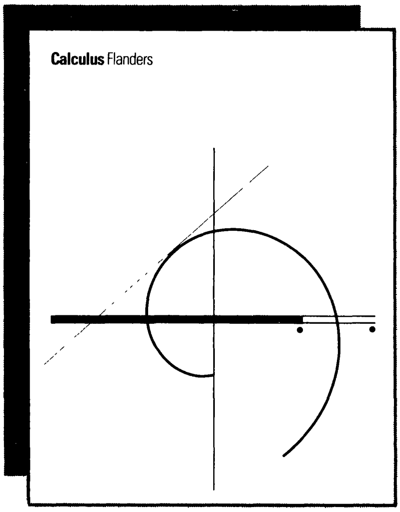
*The Mathematical Association of America gratefully acknowledges the interest and support of the following:*

CORPORATE MEMBERS

Academic Press, Inc.
Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
International Business Machines Corporation
Macmillan Publishing Company
Marcel Dekker, Inc.
The Mitre Corporation
Pergamon Press, Inc.
Prindle Weber and Schmidt, Inc.
Saunders College Publishing, Inc.
Scott Foresman & Company, Inc.
Springer-Verlag New York, Inc.
John Wiley & Sons, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories, Inc.




**HARLEY FLANDERS**

***Florida Atlantic University***
***December 1984***
***984 text pages***
***ISBN 0-7167-1643-7***

# CALCULUS

***Harley Flanders believes that the main objective of a calculus text is to help students learn to use calculus—how to set up and solve problems. And he never loses sight of that goal.***

**Efficient, economic organization**

Flanders organizes each major section by developing a small amount of introductory material on theory (leaving detailed proofs and unusual topics for optional sections), working out examples to illustrate techniques and applications, and then providing numerous and varied exercises.

**Careful, informal explanation and motivation of topics**

Major topics are introduced with informal explanations, examples, and many graphics before the mathematical details of the theory or technique are presented.

**Attention to problem solving**

Keeping the student in mind at all times, Flanders fills the text with hints and suggestions for setting up and solving problems. He even provides a unique section on problem solving to help students develop problem-solving skills early in the course.

**Avoidance of unnecessary mathematical detail**

By focusing on what is important and not being distracted by lengthy, time-consuming details, Flanders is able to develop the transcendental functions prior to integration, so that students can see some applications of integration during the first semester.

**An assurance of accuracy**

Instructors and students expect their calculus text to be accurate and error-free. The publisher has had each calculation and answer checked independently at several stages of production.

**We invite you to examine Flanders' new *Calculus* for yourself. To obtain your copy, simply address your request on school letterhead to:**

**W. H. Freeman and Company**
**41 Madison Avenue**
**New York, N.Y. 10010**

***Be sure to include your course name, current text, and anticipated enrollment.***

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 3 March 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, University of California, Berkeley, CA 94720. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# AWARD FOR DISTINGUISHED SERVICE TO PROFESSOR EVERETT PITCHER

DAVID ROSELLE
*Virginia Polytechnic Institute and State University, Blacksburg, VA 24061*

This year's recipient of the Award for Distinguished Service to Mathematics is a man who has given long, devoted and outstanding service to the American Mathematical Society and to the larger mathematics community.

Everett Pitcher was born on July 18, 1912, in Hanover, New Hampshire. His father, Arthur Dunn Pitcher, had taken a Ph.D. in general analysis under E. H. Moore at the University of Chicago in 1910 and was a faculty member at Dartmouth at the time of Everett Pitcher's birth. In 1915, the Pitcher family moved to Cleveland where Professor Pitcher served as Head of the Mathematics Department until his premature death in 1923. It is of interest that Arthur Dunn Pitcher served as a member of the Council of the AMS.

Everett Pitcher progressed through the public schools of Cleveland where his mother Wilimina Everett Pitcher also served as a teacher. He then attended Adelbert College of Western Reserve University and received the A.B. in 1932. From there, Everett proceeded to Harvard University where he received the degrees of M.A. (1933) and Ph.D. (1935).

At Harvard, Pitcher especially remembers courses in Complex Variables from William Fogg Osgood, Real Variables from Joseph L. Walsh, and Integral Equations from Marston Morse. The contact with Morse was quite enough to cause young Pitcher to ask Morse to take him on as a student. Morse was then in the midst of writing up his Colloquium Lectures and thus Pitcher was immediately given four chapters to read while still in draft. The next academic year, Pitcher and Morse worked on a problem on distribution of conjugate points and this problem became the body of Pitcher's dissertation.

When Professor Morse left Harvard University in 1935 to take up an appointment at the Institute for Advanced Study, he took the new Dr. Pitcher with him as his first assistant. Everett remembers Professor Morse as being conscious of status and thinks that he may thus have received a salary of $1800 when other assistants were paid only $1500.

Pitcher returned to Harvard in 1936 and served two years as a Benjamin Peirce Instructor. He was looking for a position of more permanence during 1937–38 and learned of just one opening, namely an instructorship at Lehigh, which he accepted.

Everett Pitcher was on active duty in the U.S. Army Ordnance Department from May 1942 until the end of 1945. His work dealt with the applications of mathematics and most of his duty was at the Ballistics Research Laboratory (BRL) working with a distinguished group of mathematicians brought together by Oswald Veblen. Colleagues included E. J. McShane, J. L. Kelley, Gerhard Hochschild, Herbert Federer, Leonard Tornheim, J. J. Levin, A. A. Bennett, Theodore Hailperin, Fritz John, J. W. Green, Frank Grubbs, Herman Goldstine, Irving Segal, I. J. Schoenberg, Hans Lewy, and Leo Zippin. The problems in which Pitcher was involved while at BRL were the accuracy of artillery firing, the terminal effects of artillery fire, and scientific intelligence; the last while on assignment in Paris and traveling in Austria during 1945.

Subsequent to his WW II assignment, Pitcher returned to the Institute for Advanced Study for the year 1945–46 where he and J. L. Kelley co-authored "Exact Sequences in Homology Theory," Ann. of Math. (2) 48, 682–709 (1947). Everett recalls this as one of his better papers and the subject matter as being as far from what he and Kelley had been doing at BRL as one could get.

Pitcher returned to Lehigh University in the fall of 1946 and was promoted twice in two years, the second time to the rank of professor. In addition to his faculty assignments, he has served Lehigh University as Department Chairman (1960–78) and as Consultant to the President (1978–present).

Everett has been enormously active in the professional activities of our community. He was Associate Secretary of the AMS (1959–66), a founder of SIAM and a member of its Board of

EVERETT PITCHER

Trustees (1961–63). He has served as Secretary of the American Mathematical Society since 1967. He was elected to Phi Beta Kappa (1930), Sigma XI (1939) and is a member of the Pi Kappa Alpha fraternity.

Everett was married to Sarah Mathiott Hindman in 1936 who died in 1972. Sarah and Everett have two children, Joan P. Morrison and Susan Pitcher-Cooper. In 1973 Everett married Theresa P. Sell, known affectionately to many of us as Terry.

Everett Pitcher has served the mathematics community exceptionally well. His service to the Society for Industrial and Applied Mathematics, the Mathematical Association of America, and to Lehigh University have been noteworthy. I feel profoundly grateful to Everett Pitcher for his exceptional service to the American Mathematical Society. I know that you share my sense of gratitude and agree with the Association's recognition of Everett Pitcher as the 1985 recipient of the Award for Distinguished Service to Mathematics.

---

## ALGORITHMIC THINKING AND MATHEMATICAL THINKING

DONALD E. KNUTH

*Computer Science Department, Stanford University, Stanford, CA 94305*

My purpose in this paper is to stimulate discussion about a philosophical question that has been on my mind for a long time: What is the actual rôle of the notion of an *algorithm* in mathematical sciences?

For many years I have been convinced that computer science is primarily the study of algorithms. My colleagues don't all agree with me, but it turns out that the source of our disagreement is simply that my definition of algorithms is much broader than theirs: I tend to think of algorithms as encompassing the whole range of concepts dealing with well-defined processes, including the structure of data that is being acted upon as well as the structure of the sequence of operations being performed; some other people think of algorithms merely as miscellaneous methods for the solution of particular problems, analogous to individual theorems in mathematics.

In the U.S.A., the sorts of things my colleagues and I do is called Computer Science, emphasizing the fact that algorithms are performed by machines. But if I lived in Germany or France, the field I work in would be called *Informatik* or *Informatique*, emphasizing the stuff that algorithms work on more than the processes themselves. In the Soviet Union, the same field is now known as either *Kibernetika* (Cybernetics), emphasizing the control of a process, or *Prikladnaia Matematika* (Applied Mathematics), emphasizing the utility of the subject and its ties to mathematics in general. I suppose the name of our discipline isn't of vital importance, since we will go on doing what we are doing no matter what it is called; after all, other disciplines like Mathematics and Chemistry are no longer related very strongly to the etymology of their names. However, if I had a chance to vote for the name of my own discipline, I would choose to call it

---

Donald E. Knuth, the Fletcher Jones Professor of Computer Science at Stanford University, is best known as the author of *The Art of Computer Programming*, a series of reference books that he plans to complete during the next two decades. His mathematical novelette, *Surreal Numbers*, has been used to help teach research methodology to undergraduates. During the past several years he has also developed new methods of mathematical typography that are now coming into wide use.

This paper has been excerpted from the text of a talk published in [**8**]. The preparation of the original version was supported in part by National Science Foundation grant MCS72-03752 A03 and in part by Office of Naval Research contract N0014-76-C-0330. This revision has been supported in part by National Science Foundation grant MCS-83-00984. Many people too numerous to name have contributed their insights during informal discussions about the topics discussed here.

Algorithmics, a word coined about 20 years ago by J. F. Traub [**12**, p. 1].

The word *algorithm* stems from the name of al-Khwârizmî, the great ninth-century scientist whose name means "from Khwârizm." When I learned the source of this word, I decided that I ought to make a pilgrimage to the Khwârizm region, since algorithms are so central to my life's work. In 1979 I had the wonderful opportunity to fulfill this dream, because a unique symposium was organized in that year by the Academy of Sciences of the Uzbek S.S.R., with the support of the Soviet Academy of Sciences. This symposium, entitled "Algorithms in Modern Mathematics and Computer Science," took place in Urgench, about 150 miles south of the Aral Sea—which was once known as Lake Khwârizm.

The chance to participate in such a symposium gave a special focus to the questions that had been buzzing around in my head. It was a time for me to think about long-range issues, things that I rarely have a chance to contemplate because of the hectic pace of activities at home. The site of the symposium, with its rich history and the grand scale of its scenery, encouraged me to look backward to the roots of my activities, and to look ahead as well—wondering what my work was all about. I asked myself several questions with renewed intensity: *What is the relation of algorithms to modern mathematics*? Is there an essential difference between an algorithmic viewpoint and the traditional mathematical world-view? *Do most mathematicians have an essentially different thinking process from that of most computer scientists*? Among members of university mathematics departments, why do the logicians (and to a lesser extent the combinatorial mathematicians) tend to be much more interested in computer science than their colleagues are?

In a sense, I had been thinking about such questions ever since my undergraduate days. I began to study higher mathematics in 1957, the same year that I began to work with digital computers; but I never mixed my mathematical thinking with my computer-science thinking in nontrivial ways until 1961. In one building I was a mathematician, in another I was a computer programmer, and it was as if I had a split personality. During 1961 I was excited by the idea that mathematics and computer science might have some common ground, because BNF notation looked mathematical; so I bought a copy of Chomsky's *Syntactic Structures* and set out to find an algorithm to decide the ambiguity problem of context-free grammars (not knowing that this had been proved impossible by Bar-Hillel, Perles, and Shamir in 1960). Needless to say, I failed to solve that problem, although I found some useful necessary and sufficient conditions for ambiguity, and I also derived a few other results like the fact that context-free languages on one letter are regular. Here, I thought, was a nice mathematical theory that I was able to develop with my computer-science intuition; how curious! During the summer of 1962, I spent a day or two analyzing the performance of hashing with linear probing, but this did not really seem like a marriage between my computer science personality and my mathematical personality since it was merely an application of combinatorial mathematics to a problem that has relevance to programming.

I think it is generally agreed that mathematicians have somewhat different thought processes from physicists, who have somewhat different thought processes from chemists, who have somewhat different thought processes from biologists. Similarly, the respective "mentalities" of lawyers, poets, playwrights, historians, linguists, farmers, and so on, seem to be unique. Each of these groups can probably recognize that other types of people have a different approach to knowledge; and it seems likely that a person gravitates to a particular kind of occupation according to the mode of thought that he or she grew up with, whenever a choice is possible. C. P. Snow wrote a famous book about "two cultures," scientific vs. humanistic, but in fact there seem to be many more than two.

Educators of computer science have repeatedly observed that only about 2 out of every 100 students enrolling in introductory programming courses really "resonate" with the subject and seem to be natural-born computer scientists. (For example, see Gruenberger [**3**].) Just before traveling to the Urgench symposium I had some independent confirmation of this, when I learned that 220 out of 11000 graduate students at the University of Illinois were majoring in Computer Science. (The relative numbers are different now, but they are probably biased by economic influences more than by natural aptitudes.) Since I believe that Computer Science is the study of

algorithms, I conclude that roughly 2% of all people "think algorithmically," in the sense that they can reason rapidly about algorithmic processes.

During my visit to the University of Illinois in 1979, I met Gerrit DeYoung, a psychologist-interested-in-computing who had recently made an interesting experiment on two groups of undergraduate students taking introductory courses in computer science. Group I consisted of 135 students intending to major in computer science, while Group II consisted of 35 social science majors. Both courses emphasized non-numeric programming and various data and control structures, although numerical problems were treated too. DeYoung handed out a questionnaire that tested each student's so-called quantitative aptitude, a standard test that seems to correlate with mathematical ability, and he also asked them to estimate their own performance in class. Afterwards he learned the grades that the students actually did receive, so he had three pieces of data on each student:

$A$ = quantitative aptitude;

$B$ = student's own perception of programming ability;

$C$ = teacher's perception of programming ability.

In both cases $B$ correlated well with $C$ (the coefficient was about .6), so we can conclude that the teachers' grading wasn't random and that there is some validity in these scores. The interesting thing was that there was *no* correlation between $A$ and $B$ or between $A$ and $C$ among the computer science majors (Group I), while there was a pronounced correlation of about .4 between the corresponding numbers for the students of Group II. It isn't clear how to interpret this data, since many different hypotheses could account for such results; perhaps psychologists know only how to measure the quantitative ability of people who think like psychologists do! At any rate the lack of correlation between quantitative ability and programming performance in the first group reminds me strongly of the feelings I often have about differences between mathematical thinking and computer-science thinking, so further study is indicated.

I believe that the real reason underlying the fact that Computer Science has become a thriving discipline at essentially all of the world's universities, although it was totally unknown twenty years ago, is *not* that computers exist in quantity; the real reason is that the algorithmic thinkers among the scientists of the world never before had a home. We are brought together in Computer Science departments *because we find people who think like we do*. At least, that seems a viable hypothesis, which hasn't been contradicted by my observations during the last half dozen or so years since the possibility occurred to me.

My goal, therefore, has been to get a deeper understanding of these phenomena; the "different modes of thought" hypothesis merely scratches the surface and gives little insight. Can we come up with a fairly clear idea of just what algorithmic thinking is, and contrast it with classical mathematical thinking?

At times when I try to come to grips with this question, I find myself almost convinced that algorithmic thinking is really like mathematical thinking, only it concentrates on more "difficult" things. But at other times I have just the opposite impression, that somehow algorithms hit only the "simpler" kinds of mathematics.... Clearly such an approach leads only to confusion and gets me nowhere.

I looked at the collection of expository works called *Mathematics: Its Content, Methods, and Meaning* [**1**] and reread what A. D. Aleksandrov had to say in his excellent introductory essay. (Interestingly enough, I found that he made prominent mention of al-Khwârizmî.) Aleksandrov listed the following characteristic features of mathematics:

- Abstractness, with many levels of abstraction.
- Precision and logical rigor.
- Quantitative relations.
- Broad range of applications.

Unfortunately, however, all four of these features seem to be characteristic also of computer science. Is there really no difference between computer science and mathematics?

**A Plan.** I decided that I could make no further progress unless I took a stab at analyzing the question "What is mathematics?"—analyzing it in some depth. Many other people have also explored this same question, and the answer is obviously that "Mathematics is what mathematicians do." More precisely, the appropriate question should probably be, "What is good mathematics?" and the answer is that "Good mathematics is what good mathematicians do."

Therefore I took nine books off of my shelf, mostly books that I had used as texts during my student days but also a few more for variety's sake. I decided to take a careful look at page 100 (i.e., a "random" page) in each book and to study the first result on that page. This way I could get a sample of what good mathematicians do, and I could attempt to understand the types of thinking that seem to be involved.

From the standpoint of computer science, the notion of "types of thinking" is not so vague as it once was, since we can now imagine trying to make a computer program discover the mathematics. What sorts of capabilities would we have to put into such an artificially intelligent program, if it were to be able to come up with the results on page 100 of the books I selected?

In order to make this experiment fair, I was careful to abide by the following ground rules: (1) The books were all to be chosen first, before I studied any particular one of them. (2) Page 100 was to be the page examined in each case, since I had no *a priori* knowledge of what was on that page in any book. If somehow page 100 turned out to be a bad choice, I wouldn't try anything sneaky like searching for another page number that would give results more in accord with my prejudices. (3) I would not suppress any of the data; every book I had chosen would appear in the final sample, so that I wouldn't introduce any bias by selecting a subset.

The results of this experiment opened up my eyes somewhat, so I would like to share them with you. Here is a book-by-book summary of what I found.

**Book 1: Thomas's Calculus.** I looked first at the book that had introduced me to higher mathematics, the calculus text by George B. Thomas [**11**] that I had used as a college freshman. On page 100 he treats the following problem: *What value of* $x$ *minimizes the travel time from* $(0, a)$ *to* $(x, 0)$ *to* $(d, -b)$, *if you must go at speed* $s_1$ *from* $(0, a)$ *to* $(x, 0)$ *and at some other speed* $s_2$ *from* $(x, 0)$ *to* $(d, -b)$?

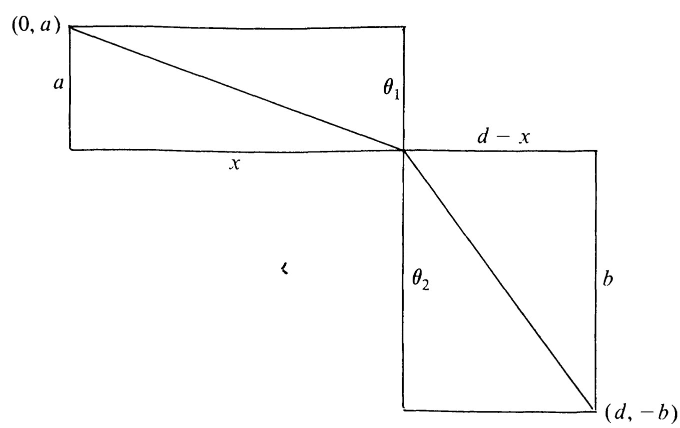



FIG. 1

In other words, we want to minimize the function

$$f(x) = \sqrt{a^2 + x^2}\,/s_1 + \sqrt{b^2 + (d - x)^2}\,/s_2.$$

The solution is to differentiate $f(x)$, obtaining

$$f'(x) = \frac{x}{s_1\sqrt{a^2+x^2}} - \frac{d-x}{s_2\sqrt{b^2+(d-x)^2}} = \frac{\sin\theta_1}{s_1} - \frac{\sin\theta_2}{s_2}.$$

As $x$ runs from 0 to $d$, the value of $(\sin\theta_1)/s_1$ starts at zero and increases, while the value of $(\sin\theta_2)/s_2$ decreases to zero. Therefore the derivative starts negative and ends positive; there must be a point where it is zero, i.e., $(\sin\theta_1)/s_1 = (\sin\theta_2)/s_2$, and that's where the minimum occurs. Thomas remarks that this is "Snell's Law" in optics; somehow light rays know how to minimize their travel time.

The mathematics involved here seems to be mostly a systematic procedure for minimization, based on formula manipulation and the correspondence between formulas and geometric figures, together with some reasoning about changes in function values. Let us keep this in mind as we look at the other examples, to see how much the examples have in common.

**Book 2: A Survey of Mathematics.** Returning to the survey volumes edited by Aleksandrov et al. [**1**], we find that page 100 is the chapter on Analysis by Lavrent'ev and Nikol'skiĭ. It shows how to deduce the derivative of the function $\log_a x$ in a clever way:

$$\frac{\log_a(x+h) - \log_a x}{h} = \frac{1}{h}\log_a\frac{x+h}{x} = \frac{1}{x}\log_a\left(1+\frac{h}{x}\right)^{x/h}.$$

The logarithm function is continuous, so we have

$$\lim_{h\to 0}\frac{1}{x}\log_a\left(1+\frac{h}{x}\right)^{x/h} = \frac{1}{x}\log_a\lim_{h\to 0}\left(1+\frac{h}{x}\right)^{x/h} = \frac{1}{x}\log_a e,$$

since it has already been proved that the quantity $(1+1/n)^n$ approaches a constant called $e$, when $n$ approaches infinity through integer or noninteger values. Here the reasoning involves formula manipulation and an understanding of limiting processes.

**Book 3: Kelley's General Topology.** The third book I chose was a standard topology text [**4**], where page 100 contains the following exercise: "*Problem* A. *The image under a continuous map of a connected space is connected.*" No solution is given, but I imagine something like the following was intended: First we recall the relevant definitions, that a function $f$ from topological space $X$ to topological space $Y$ is continuous when the inverse image $f^{-1}(V)$ is open in $X$, for all open sets $V$ in $Y$; a topological space $X$ is connected when it cannot be written as a union of two nonempty open sets. Thus, let us try to prove that $Y$ is connected, under the assumption that $f$ is continuous and $X$ is connected, where $f(X) = Y$. If $Y = V_1 \cup V_2$, where $V_1$ and $V_2$ are disjoint and open, then $X = f^{-1}(V_1) \cup f^{-1}(V_2)$, where $f^{-1}(V_1)$ and $f^{-1}(V_2)$ are disjoint and open. It follows that either $f^{-1}(V_1)$ or $f^{-1}(V_2)$ is empty, say $f^{-1}(V_1)$ is empty. Finally, therefore, $V_1$ is empty, since $V_1 \subset f(f^{-1}(V_1))$. Q.E.D.

(Note that no properties of "open sets" were needed in this proof.)

The mathematical thinking involved here is somewhat different from what we have seen before; it consists primarily of constructing chains of implications from the hypotheses to the desired conclusions, using a repertoire of facts like "$f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$". This is analogous to constructing chains of computer instructions that transform some input into some desired output, using a repertoire of subroutines, although the topological facts have a more abstract character.

Another type of mathematical thinking is involved here, too, and we should be careful not to forget it: Somebody had to define the concepts of continuity and connectedness in some way that would lead to a rich theory having lots of applications, thereby generalizing many special cases that had been proved before the abstract pattern was perceived.

**Book 4: From the 18th Century.** Another book on my list was Struik's *Source Book in Mathematics*, which quotes authors of famous papers written during the period 1200–1800 A.D. Page 100 is concerned with Euler's attempt to prove the fundamental theorem of algebra, in the

course of which he derived the following auxiliary result: "*Theorem* 4. *Every quartic polynomial* $x^4 + Ax^3 + Bx^2 + Cx + D$ *with real coefficients can be factored into two quadratics.*"

Here's how he did it. First he reduced the problem to the case $A = 0$ by setting $x = y - \frac{1}{4}A$. Then he was left with the problem of solving $(x^2 + ux + \alpha)(x^2 - ux + \beta) = x^4 + Bx^2 + Cx + D$ for $u$, $\alpha$, and $\beta$, so he wanted to solve the equations $B = \alpha + \beta - u^2$, $C = (\beta - \alpha)u$, $D = \alpha\beta$. These equations lead to the relations $2\beta = B + u^2 + C/u$, $2\alpha = B + u^2 - C/u$, and $(B + u^2)^2 - C^2/u^2 = 4D$. But the cubic polynomial $(u^2)^3 + 2B(u^2)^2 + (B^2 - 4D)u^2 - C^2$ goes from $-C^2$ to $+\infty$ as $u^2$ runs from 0 to $\infty$, so it has a positive root, and the factorization is complete.

(Euler went on to generalize, arguing that every polynomial of degree $2^n$ can be factored into two of degree $2^{n-1}$, via a polynomial of odd degree $\frac{1}{2}\binom{2^n}{2^{n-1}}$ in $u^2$ having a negative constant term. But this part of his derivation was not rigorous; Lagrange and Gauss later pointed out a serious flaw.)

When I first looked at this example, it seemed to be distinctly more "algorithmic" than the preceding ones—probably because Euler was essentially explaining how to take a quartic polynomial as input and to produce two quadratic polynomials as output. Input/output characteristics are significant aspects of algorithms, although Euler's actual construction is comparatively simple and direct so it doesn't exhibit the complex control structure that algorithms usually have. The types of thinking involved here seem to be (a) to reduce a general problem to a simpler special case (by showing that $A$ can be assumed zero, and by realizing that the resulting sixth-degree equation in $u$ is really a third-degree equation in $u^2$); (b) formula manipulation to solve simultaneous equations for $\alpha$, $\beta$, and $u$; (c) generalization by recognizing a pattern for the case of 4th degree equations that apparently would extend to degrees 8, 16, etc.

**Book 5: Abstract Algebra.** My next choice was another standard textbook, *Commutative Algebra* by Zariski and Samuel [**14**]. Their page 100 is concerned with the general structure of arbitrary fields. Suppose $k$ and $K$ are fields with $k \subset K$; the *transcendence degree of $K$ over $k$* is defined to be the cardinal number of any "transcendence basis" $L$ of $K$ over $k$, namely a set $L$ such that all of its finite subsets are algebraically independent over $k$ and such that all elements of $K$ are algebraic over $k(L)$; i.e., they are roots of polynomial equations whose coefficients are in the smallest field containing $k \cup L$. The exposition in the book has just found that this cardinal number is a well-defined invariant of $k$ and $K$, i.e., that all transcendence bases of $K$ over $k$ have the same cardinality.

Now comes Theorem 26: *If $k \subset K \subset \mathscr{K}$, the transcendence degree of $\mathscr{K}$ over $k$ is the sum of the transcendence degrees of $K$ over $k$ and of $\mathscr{K}$ over $K$.* To prove the theorem, Zariski and Samuel let $L$ be a transcendence basis of $K$ over $k$ and $\mathscr{L}$ a transcendence basis of $\mathscr{K}$ over $K$; the idea is to prove that $L \cup \mathscr{L}$ is a transcendence basis of $\mathscr{K}$ over $k$, and the result follows since $L$ and $\mathscr{L}$ are disjoint.

The required proof is not difficult and it is worth studying in detail. Let $\{x_1, \ldots, x_m, X_1, \ldots, X_M\}$ be a finite subset of $L \cup \mathscr{L}$, where the $x$'s are in $L$ and the $X$'s in $\mathscr{L}$, and assume that they satisfy some polynomial equation over $k$, namely

$$(*) \qquad \sum_{\substack{e_1, \ldots, e_m \geq 0 \\ E_1, \ldots, E_M \geq 0}} \alpha(e_1, \ldots, e_m, E_1, \ldots, E_M)\, x_1^{e_1} \cdots x_m^{e_m} X_1^{E_1} \cdots X_M^{E_M} = 0,$$

where all the $\alpha(e_1, \ldots, e_m, E_1, \ldots, E_M)$ are in $k$ and only finitely many $\alpha$'s are nonzero. This equation can be rewritten as

$$(**) \qquad \sum_{E_1, \ldots, E_M \geq 0} \left( \sum_{e_1, \ldots, e_m \geq 0} \alpha(e_1, \ldots, e_m, E_1, \ldots, E_M)\, x_1^{e_1} \cdots x_m^{e_m} \right) X_1^{E_1} \cdots X_M^{E_M} = 0,$$

a polynomial in the $X$'s with coefficients in $K$, hence all of these coefficients are zero by the algebraic independence over $\mathscr{L}$ over $K$. These coefficients in turn are polynomials in the $x$'s with coefficients in $k$, so all the $\alpha$'s must be zero. In other words, any finite subset of $L \cup \mathscr{L}$ is

algebraically independent.

Finally, all elements of $K$ are algebraic over $k(L)$ and all elements of $\mathscr{K}$ are algebraic over $K(\mathscr{L})$. It follows from the previously developed theory of algebraic extensions that all elements of $\mathscr{K}$ are algebraic over $k(L)(\mathscr{L})$, the smallest field containing $k \cup L \cup \mathscr{L}$. Hence $L \cup \mathscr{L}$ satisfies all the criteria of a transcendence basis.

Note that the proof involves somewhat sophisticated "data structures," i.e., representations of complex objects, in this case polynomials in many variables. The key idea is a *pun*, the equivalence between the polynomial over $k$ in $(*)$ and the polynomial over $k(L)$ in $(**)$. In fact, the structure theory of fields being developed in this part of Zariski and Samuel's book is essentially a theory about data structures by which all elements of the field can be represented and manipulated. Theorem 26 is not as important as the construction of transcendence bases that appears in its proof.

Another noteworthy aspect of this example is the way infinite sets are treated. Finite concepts have been generalized to infinite ones by saying that all finite subsets must have the property; this allows algorithmic constructions to be applied to the subsets.

**Book 6: Metamathematics.** I chose Kleene's *Introduction to Metamathematics* [5] as a representative book on logic. Page 100 talks about "disjunction elimination": Suppose we are given (1) $\vdash A \vee B$ and (2) $A \vdash C$ and (3) $B \vdash C$. Then by a rule that has just been proved, (2) and (3) yield

$$A \vee B \vdash C. \tag{4}$$

From (1) and (4) we may now conclude "(5) $\vdash C$". Kleene points out that this is the familiar idea of reasoning by cases. If either $A$ or $B$ is true, we can consider case 1 that $A$ is true (then $C$ holds); or case 2 that $B$ is true (and again $C$ holds). It follows that statement $C$ holds in any case.

The reasoning in this example is simple formula manipulation, together with an understanding that familiar thought patterns are being generalized and made formal.

I was hoping to hit a more inherently metamathematical argument here, something like "anything that can be proved in system $X$ can also be proved in system $Y$," since such arguments are often essentially algorithms that convert arbitrary $X$-proofs into $Y$-proofs. But page 100 was more elementary, this being an introductory book.

**Book 7: Knuth.** Is my own work [6] algorithmic? Well, page 100 isn't especially so, since it is part of the introduction to mathematical techniques that appear before I get into the real computer science content. The problem discussed on that page is to get the mean and standard deviation of the number of "heads" in $n$ coin flips, when each independent flip comes up "heads" with probability $p$ and "tails" with probability $q = 1 - p$. I introduce the notation $p_{nk}$ for the probability that $k$ heads occur, and observe that

$$p_{nk} = p \cdot p_{n-1,k-1} + q \cdot p_{n-1,k}.$$

To solve this recurrence, I introduce the generating function

$$G_n(z) = \sum_{k \geq 0} p_{nk} z^k$$

and obtain $G_n(z) = (q + pz)G_{n-1}(z)$, $G_1(z) = q + pz$. Hence $G_n(z) = (q + pz)^n$, and

$$\operatorname{mean}(G_n) = n \operatorname{mean}(G_1) = pn; \qquad \operatorname{var}(G_n) = n \operatorname{var}(G_1) = pqn.$$

Thus, the recurrence relation is set up by reasoning about probabilities; it is solved by formula manipulation according to patterns that are discussed earlier in the book.

**Book 8: Pólya and Szegő.** The good old days of mathematics are represented by Pólya and Szegő's famous *Aufgaben und Lehrsätze*, recently available in an English translation with many new Aufgaben [9]. Page 100 contains a real challenge:

**217.** $$\lim_{n\to\infty} \int_{-\pi}^{\pi} \frac{n!\,2^{2n\cos\theta}}{|(2ne^{i\theta} - 1)\cdots(2ne^{i\theta} - n)|}\,d\theta = 2\pi.$$

Fortunately the answer pages provide enough of a clue to reveal the proof that the authors had in mind. We have

$$|2ne^{i\theta} - k|^2 = 4n^2 + k^2 - 4nk\cos\theta = (2n-k)^2 + 4nk(1-\cos\theta)$$
$$= (2n-k)^2 + 8nk\sin^2\theta/2.$$

Replacing $\theta$ by $x/\sqrt{n}$ allows us to rewrite the integral as

$$\frac{n!\,2^{2n}}{((2n-1)\cdots n)\sqrt{n}}\int_{-\infty}^{\infty} f_n(x)\,dx,$$

where $f_n(x) = 0$ for $|x| > \pi\sqrt{n}$, and otherwise

$$f_n(x) = 2^{2n(\cos x/\sqrt{n}-1)}\prod_{1\leqslant k\leqslant n}\left(1 + \frac{8nk}{(2n-k)^2}\sin^2\frac{x}{2\sqrt{n}}\right)^{-1/2}$$
$$= \exp\left((2\ln 2)n\left(\cos\frac{x}{\sqrt{n}} - 1\right) - \sum_{1\leqslant k\leqslant n}\frac{1}{2}\ln\left(1 + \frac{8nk}{(2n-k)^2}\sin^2\frac{x}{2\sqrt{n}}\right)\right)$$
$$= \exp\left(-x^2\ln 2 + O\left(\frac{x^4}{n}\right) + \frac{1}{2}\sum_{1\leqslant k\leqslant n}\left(\frac{-2nk}{(2n-k)^2}\frac{x^2}{n} + O\left(\frac{x^4}{n^2}\right)\right)\right)$$
$$= \exp\left(-x^2\ln 2 - (1-\ln 2)x^2 + O\left(\frac{1+x^4}{n}\right)\right).$$

Thus, $f_n(x)$ converges uniformly to $e^{-x^2}$ in any bounded interval. Furthermore we have $|f_n(x)| \leqslant 2^{2n(\cos x/\sqrt{n}-1)}$ and

$$\cos\frac{x}{\sqrt{n}} - 1 \leqslant -\frac{x^2}{2n} + \frac{x^4}{24n^2} \leqslant -\left(\frac{1}{2} - \frac{\pi^2}{24}\right)\frac{x^2}{n} \text{ for } |x| \leqslant \pi/\sqrt{n},$$

since the cosine function is "enveloped" by its Maclaurin series; therefore $|f_n(x)|$ is less than the integrable function $e^{-cx^2}$ for all $n$, where $c = 1 - \pi^2/12$. From this uniformly bounded convergence we are justified in taking limits past the integral sign,

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

Finally, the coefficient in front of $\int_{-\infty}^{\infty} f_n(x)\,dx$ is $2^{2n+1}n!^2/\sqrt{n}(2n)!$, which is equal to $2\sqrt{\pi}(1 + O(1/n))$ by Stirling's approximation, and the result follows.

This derivation gives some idea of how far mathematics had developed between the time of al-Khwârizmî and 1920. It involves formula manipulation and an understanding of the asymptotic limiting behavior of functions, together with the idea of inventing a suitable function $f_n$ that will rigorously permit us to make the interchange $\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty}(\lim_{n\to\infty} f_n(x))\,dx$. The definition of $f_n(x)$ requires a clear understanding of how functions like $\exp x$ and $\cos x$ behave.

**Book 9: Bishop's Constructive Mathematics.** The last book I chose to sample turned out to be the most interesting of all from the standpoint of my quest; it was Errett Bishop's *Foundations of Constructive Mathematics* [2], a book that I had heard about but never before read. The interesting thing about this book is that it reads essentially like ordinary mathematics, yet it is entirely algorithmic in nature if you look between the lines.

Page 100 of Bishop's book contains Corollary 3 to the Stone-Weierstrass theorem developed on the preceding pages: *Every uniformly continuous function on a compact set $X \subset$* **R** *can be arbitrarily closely approximated on $X$ by polynomial functions over* **R**. And here is his proof: "By Lemma 5, the function $x \mapsto |x - x_0|$ can be arbitrarily closely approximated on $X$ by polynomials. The theorem then follows from Corollary 2."

We might call this a compact proof! Before unwrapping it to explain what Lemma 5 and Corollary 2 are, I want to stress that the proof is essentially an algorithm; the algorithm takes any constructively given compact set $X$ and continuous function $f$ and tolerance $\varepsilon$ as input, and it outputs a polynomial that approximates $f$ to within $\varepsilon$ on all points of $X$. Furthermore the algorithm operates on algorithms, since $f$ is given by an algorithm of a certain type, and since real numbers are essentially algorithms themselves.

I will try to put Bishop's implicit algorithms into an explicit PASCAL-like form, even though the capabilities of today's programming languages have to be stretched considerably to reflect his constructions. First let's consider Lemma 5, which states that for each $\varepsilon > 0$ there exists a polynomial $p\colon \mathbf{R} \to \mathbf{R}$ such that $p(0) = 0$ and $||x| - p(x)| \leqslant \varepsilon$ for all $|x| \leqslant 1$. Bishop's proof, which makes the lemma an algorithm, is essentially the following.

> **function** *Lemma* 5 ($\varepsilon$: *real*): **R**-*polynomial*;
> **var** $N$: *integer*; $g, p$: **R**-*polynomial*;
> **begin** $N$ := suitable function of $\varepsilon$;
> $g(t) := 1 - \Sigma_{1 \leqslant n \leqslant N}\binom{1/2}{n}(-1)^n t^n; \qquad p(t) := g(-t^2) - g(1);$
> *Lemma* 5 := $p$;
> **end**.

Here $N$ is to be computed large enough that $|g(t) - (1-t)^{1/2}| \leqslant \frac{1}{2}\varepsilon$ for $0 \leqslant t \leqslant 1$.

The other missing component of the proof on page 100 is Corollary 2, which states that if $X$ is any compact metric space and if $G$ is the set of all functions $x \mapsto \rho(x, x_0)$, where $x_0 \in X$ and where $\rho(x, y)$ denotes the metric distance from $x$ to $y$, then "$\mathscr{A}(G)$ is dense in $C(X)$." That is, all uniformly continuous real-valued functions on $X$ can be approximated to arbitrarily high accuracy by functions obtained from the functions $G$ by a finite number of operations of addition, multiplication, and multiplication by real numbers. Bishop's Corollary 2 turns out to be false in the case that $X$ contains only one point, since $G$ and $\mathscr{A}(G)$ then consist only of the zero function. (I noticed this oversight while trying to formulate his proof in an explicitly algorithmic way.) But the defect is easily remedied.

For our purposes it is best to reformulate Bishop's Corollary 2 as follows: "*Let $X$ be a compact metric space containing at least two points, and let $G$ be the set of all functions of the form $x \mapsto c\rho(x, x_0)$, where $c > 0$ and $x_0 \in X$. Then $G$ is a separating family over $X$.*" I'll repeat Bishop's definition of separating family in a minute; first I want to mention his Theorem 7, the Stone-Weierstrass theorem whose proof I shall not discuss in detail, namely the fact that $\mathscr{A}(G)$ is dense in $C(X)$ whenever $G$ is a separating family of uniformly continuous functions over a compact metric space $X$. In view of this theorem, my reformulation of Corollary 2 leads to the corollary as he stated it.

A *separating family* is a collection of real-valued functions $G$ over $X$, together with a function $\delta$ from the positive reals $\mathbf{R}^+$ into $\mathbf{R}^+$, and also together with two selection algorithms $\sigma$ and $\tau$. Algorithm $\sigma$ takes elements $x, y$ of $X$ and a positive real number $\varepsilon$ as input, where $\rho(x, y) \geqslant \varepsilon$, and selects an element $g$ of $G$ such that for all $z$ in $X$ we have

$$\rho(x, z) \leqslant \delta(\varepsilon) \text{ implies } |g(z)| \leqslant \varepsilon, \qquad \rho(y, z) \leqslant \delta(\varepsilon) \text{ implies } |g(z) - 1| \leqslant \varepsilon.$$

Algorithm $\tau$ takes an element $y$ of $X$ and a positive real number $\varepsilon$ as input, and selects an element $g$ of $G$ such that the second of the implications above holds, for all $z$ in $X$.

Thus the reformulated Corollary 2 is an algorithm that takes a nontrivial compact metric space $X$ as input and yields a separating family $(\delta, \sigma, \tau)$, where $\sigma$ and $\tau$ select functions of the form $\rho(x, x_0)$. Here is the construction:

> **function** *Corollary* 2($X$: *compact metric space*;
> $y_0, y_1$: *X-element*): *X-separating family*;
> $\{y_0$ and $y_1$ are distinct elements of $X\}$
> **var** $\delta$: **function** $\mathbf{R}^+ \to \mathbf{R}^+$;

```
    d: function X × X → R⁺;
    σ: function X × X × R⁺ → C(X);
    τ: function X × R⁺ → C(X);
begin d(x, y) := X. ρ(x, y); {this is the distance function in X}
δ(ε) := min(ε², ½εd(y₀, y₁)));
σ(x, y, ε) := (function g(z: X-element): R;
                         g := d(x, z)/d(x, y));
τ(y, ε) := (function g(z: X-element): R;
                      g := (if d(y, y₁) ≤ ½d(y₀, y₁)
                            then d(y, z)/d(y, y₀)
                            else d(y, z)/d(y, y₁)))
Corollary 2 := (δ, σ, τ);
end.
```

My notation for the complicated types involved in these algorithms is not the best possible, but I hope it is reasonably comprehensible without further explanation. The selection rule $\sigma$ determined by this algorithm has the desired property since, for example, $\rho(x,y) > \varepsilon$ and $\rho(y,z) \leqslant \delta(\varepsilon) \leqslant \varepsilon^2$ implies that $|g(z) - 1| = |\rho(x,z) - \rho(x,y)|/\rho(x,y) \leqslant \rho(y,z)/\rho(x,y) \leqslant \varepsilon$.

Bishop's proof of Corollary 3 can now be displayed more explicitly as an algorithm in the following way. If $X$ is a compact subset of $\mathbf{R}$, under Bishop's definition, we can compute $M = \text{bound}(X)$ such that $X$ is contained in the closed interval $[-M, M]$. Let us assume that his Theorem 7 is a procedure whose input parameters consist of a compact metric space $X$, a separating family $(\delta, \sigma, \tau)$ over $X$ that selects functions from some set $G \subset C(X)$, and a uniformly continuous function $f\colon X \to \mathbf{R}$, and a positive real number $\varepsilon$. The output of this procedure is an element $A$ of $\mathscr{A}(G)$, namely a finite sum of terms of the form $Cg_1(x)\dots g_m(x)$ where $m \geqslant 1$ and each $g_i \in G$; this output satisfies $|A(x) - f(x)| \leqslant \varepsilon$ for all $x$ in $X$.

Here is the fleshed-out form of Corollary 3:

```
function Corollary 3 (X: compact real set;
                      f: X-continuous function;
                      ε: positive): R-polynomial;
var p, q, r: R-polynomial; M, B: real; y₀, y₁: X-element;
  A: 𝒜(G)-element, where G is the set of functions x ↦ c|x − x₀|;
begin M := bound (X);
y₀ := element (X);
if trivial (X) then r(t) := f(y₀)
else begin y₁ := element (X \ {y₀});
    A := Theorem 7 (X, Corollary 2 (X, y₀, y₁), f, ½ε);
    B := suitable function of A, see below;
  p(t) := Lemma 5 (ε/B);
  q(t) := 2Mp(t/2M);
      {||x − x₀| − q(x − x₀)| ≤ ε/B for all x}
  r(x) := substitute cq(x − x₀) for each factor gᵢ(x) = c|x − x₀|
          of each term of A;
    {B was chosen so that |q(x − x₀) − |x − x₀|| ≤ ε/B
              implies that |r(x) − A(x)| ≤ ½ε}
  end;
Corollary 3 := r;
end.
```

Clearly it would be an extremely interesting project from the standpoint of high-level programming language design to find an elegant notation in which Bishop's constructions are both readable and explicit.

**Tentative Conclusions.** What insights do we get from these nine randomly selected examples of mathematics? In the first place, they point out something that should have been obvious to me from the start, that there is no such thing as "mathematical thinking" as a single isolated concept; mathematicians use a variety of modes of thought, not just one. My question about computer-science thinking as distinct from math thinking therefore needs to be reformulated. Indeed, when I reflect further about my student days, I realize that I would not only wear my CS hat when programming computers and my math hat when taking courses, I also had other hats representing various modes of thought that I used when I was editing a student magazine or when I was acting as officer of a fraternity, etc.

Thus, it seems better to think of a model in which people have a certain number of different modes of thought, something like genes in DNA. It is probable that computer scientists and mathematicians overlap in the sense that they share several modes of thought, yet there are other modes peculiar to one or the other. Under this model, different areas of science can be characterized by different "personality profiles."

After trying to distill out different kinds of reasoning in the nine examples, I came up with nine categories that I tentatively would diagram as follows. (Two $x$'s means a strong use of some reasoning mode, while one $x$ indicates a mild connection.)

| | Formula manipulation | Representation of reality | Behavior of function values | Reduction to simpler problems | Dealing with infinity | Generalization | Abstract reasoning | Information structures | Algorithms |
|---|---|---|---|---|---|---|---|---|---|
| 1 (Thomas) | $xx$ | $xx$ | $xx$ | | | | | | |
| 2 (Lavrent'ev) | $xx$ | | $x$ | | $xx$ | | | | |
| 3 (Kelley) | $x$ | | | | | $xx$ | $xx$ | | |
| 4 (Euler) | $xx$ | | $xx$ | $x$ | | $xx$ | | | $x$ |
| 5 (Zariski) | $x$ | | | $x$ | $xx$ | $x$ | $xx$ | $xx$ | |
| 6 (Kleene) | $x$ | | | | | $xx$ | $xx$ | | $x$ |
| 7 (Knuth) | $xx$ | $x$ | | $x$ | | | | | |
| 8 (Pólya) | $xx$ | | $xx$ | $xx$ | $xx$ | | | | |
| 9 (Bishop) | $xx$ | | $xx$ | $xx$ | | $x$ | $xx$ | $xx$ | $x$ |
| "Algorithmic thinking" | $x$ | $xx$ | | $xx$ | | | $xx$ | $xx$ | $xx$ |

These nine categories aren't precisely defined, and they may represent combinations of more fundamental things; for example, both formula manipulation and generalization involve the general idea of pattern recognition (spotting certain kinds of order). Another fundamental distinction might be in the type of "visualization" needed, whether it be geometric or abstract or recursive, etc. Thus, I am not at all certain of the categories; they are simply put forward as a basis for discussion.

I have added a tenth row to the table labeled "algorithmic thinking," trying to make it represent my perception of the most typical thought processes used by a computer scientist. Since computer science is such a young discipline, I don't know what books would be appropriate candidates from which to examine page 100. It seems to me that most of the modes of thought listed in the table are common in computer science as well as in mathematics, with the notable exception of "dealing with infinity." Infinite-dimensional spaces seem to be of little relevance for computer scientists, although most other branches of mathematics have been extensively applied in many different ways.

Computer scientists will notice, I think, that two types of thinking are absent from the examples we have studied, so this may be what separates mathematicians from computer scientists. In the first place, there is almost no notion of "complexity" or economy of operation in what we have discussed. Bishop's mathematics is constructive, but it does not have all the ingredients of an algorithm because it ignores the "cost" of the constructions. If we carry out the details of his Stone-Weierstrass theorem with respect to simple functions, we are likely to wind up with a polynomial approximation of degree $10^6$, say, although a suitable polynomial of degree 6 could have been found by a more efficient scheme.

The other missing concept is related to the "assignment operation" $:=$ , which changes values of quantities. More precisely, I would say that the missing concept is the dynamic notion of the *state* of a process: "How did I get here? What is true now? What should happen next if I'm going to get to the end?" Changing states of affairs, or snapshots of a computation, seem to be intimately related to algorithms and algorithmic thinking. Many of the concepts of data structures, which are so fundamental in computer science, depend very heavily on an ability to reason about the notion of process states, and we rely on this notion also when studying the interaction of processes that are acting simultaneously.

Our nine examples don't have anything resembling "$n := n + 1$", except for Euler's discussion where he essentially begins by setting $x := x - \frac{1}{4}A$. The assignment operations in Bishop's constructions aren't really assignments, they are simply definitions of quantities, and those definitions won't be changed. This discrepancy between classical mathematics and computer science is well illustrated by the fact that Burks, Goldstine, and von Neumann did not actually have the notion of assignment in their early notes on computer programming; they used a curious in-between concept instead (see **[7]**).

A computer scientist also seems to be much more willing to deal with a multitude of quite different cases than does a traditional mathematician. Data structures in computer science needn't be homogeneous, and algorithms can involve many different kinds of steps. Sometimes that is a weakness of computer scientists, because we don't try as hard as we should to find uniformity; but sometimes it is a strength because we can deal fluently with concepts that are inherently non-uniform.

At the Urgench symposium I had the pleasure of hearing a stimulating lecture by G. S. Tseytin about these issues; I heartily recommend his paper **[13]**.

### References

**1.** A. D. Aleksandrov, A. N. Kolmogorov, and M. A. Lavrent'ev, eds., Mathematics: Its Content, Methods and Meaning **1** (Cambridge, Mass.: MIT Press, 1963). Translated by S. H. Gould and T. Bartha from Matematika: Eë Soderzhanie, Metody i Znachenie (Akad. Nauk SSSR, 1956).

**2.** Errett Bishop, Foundations of Constructive Analysis (N.Y.: McGraw-Hill, 1967).

**3.** Fred Gruenberger, "The role of education in preparing effective computing personnel," in F. Gruenberger, ed., Effective vs. Efficient Computing (Englewood Cliffs, N.J.: Prentice-Hall, 1973), 112–120.

**4.** John L. Kelley, General Topology (Princeton: D. Van Nostrand, 1955).

**5.** Stephen Cole Kleene, Introduction to Metamathematics (Princeton: D. Van Nostrand, 1952).

**6.** Donald E. Knuth, Fundamental Algorithms (Reading, Mass.: Addison-Wesley, 1968).

**7.** Donald E. Knuth and Luis Trabb Pardo, "The early development of programming languages," Encyclopedia of Computer Science and Technology **7** (N.Y.: Marcel Dekker, 1977), 419–493.

**8.** Donald E. Knuth, "Algorithms in modern mathematics and computer science," Lecture Notes in Computer Science **122** (Berlin: Springer, 1981), 82–99.

**9.** G. Pólya and G. Szegő, Problems and Theorems in Analysis **1** (Berlin: Springer, 1972).

**10.** D. J. Struik, ed., A Source Book in Mathematics, 1200–1800 (Cambridge, Mass.: Harvard University Press, 1969).

**11.** George B. Thomas, Calculus and Analytic Geometry, 2nd ed. (Cambridge, Mass.: Addison-Wesley, 1956).

**12.** J. F. Traub, Iterative Methods for the Solution of Equations (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

**13.** G. S. Tseytin, "From logicism to proceduralism (An autobiographical account)," Lecture Notes in Computer Science **122** (Berlin: Springer, 1981), 390–396.

**14.** Oscar Zariski and Pierre Samuel, Commutative Algebra **1** (Princeton: D. Van Nostrand, 1958).

They are more closely related than most pairs of mathematicians. (See p. 214.)

tedious drill and leave the teacher free to discuss the important conceptual aspects of integration. I take the consensus to be that change of variable, integration by parts, and the use of partial fractions must be taught: change of variable because of its theoretical importance and because it gives the student a chance to see how a hard problem becomes easy when looked at in the right way; integration by parts because of its constant use in both pure and applied mathematics as a theoretical tool; and partial fractions because one can show how only two transcendental functions (the logarithm and the arctangent) are necessary for integrating all rational functions.

Having agreed to this, what problems do we discuss to make our points? It seems to me they should be conceptual problems that the student can appreciate. I offer one example that I have found useful.

The students know how to compute the area, $\pi r^2$, and the perimeter, $2\pi r$, of a circle of radius $r$. The good students may even have wondered why this same mysterious $\pi$ appears in both formulae. For those who have not, this question gets their attention.

Assuming the students know how to use the definite integral to find area and arclength, one has

$$A = 4\int_0^r \sqrt{r^2 - x^2}\, dx, \quad P = 4\int_0^r \frac{r}{\sqrt{r^2 - x^2}}\, dx.$$

Now, the change of variable, $x = ru$, gives immediately that

$$A = \left\{4\int_0^1 \sqrt{1 - x^2}\, dx\right\} r^2$$

and

$$P = 2\left\{2\int_0^1 \frac{1}{\sqrt{1 - x^2}}\, dx\right\} r.$$

So, it is a question of relating

$$\int_0^1 \sqrt{1 - x^2}\, dx \text{ to } \int_0^1 \frac{1}{\sqrt{1 - x^2}}\, dx.$$

But the identity $1 = (1 - x^2) + x^2$ and integration by parts yield

$$\int_0^1 \frac{1}{\sqrt{1 - x^2}}\, dx = 2\int_0^1 \sqrt{1 - x^2}\, dx.$$

We've actually proved the existence of a number, $\pi$ say, with $A = \pi r^2$ and $P = 2\pi r$; we can calculate its approximate value by numerical integration; and we've done so using change of variable and integration by parts. Who could ask for anything more?

---

## 142. MISCELLANEA

Beauty is perfect, and perfection (such is human nature) holds our attention but for a little while. The mathematician who after seeing Phèdre asked: "*Qu'est-ce que ça prouve*?" was not such a fool as he has been generally made out.

—W. Somerset Maugham, *Cakes and Ale*.

---

**ANSWER TO PHOTOS ON PAGE 182**

They are the brothers Alfred and Richard Brauer.

# THE ANATOMY OF LOW DIMENSIONAL STABLE SINGULARITIES

JONATHAN BRITT

*Department of Mathematics and Computing, Southampton College of Higher Education, Southampton S09 4WW, England*

**Preface.** In this article we examine the geometry of certain sets, known as singularity sets, associated with a particular mapping called the Whitney cusp mapping. We will also be looking at the geometrical structure of the inverse image of those sets under that mapping.

Initially we will consider the mapping from $\mathbb{R}^2$ to itself and then from $\mathbb{R}^3$ to $\mathbb{R}^2$. We will then examine the most general case of the Whitney cusp mapping that is from $\mathbb{R}^n$ to $\mathbb{R}^2$. Finally we will explore, in the same way, the structure of the swallowtail mapping from $\mathbb{R}^3$ to itself.

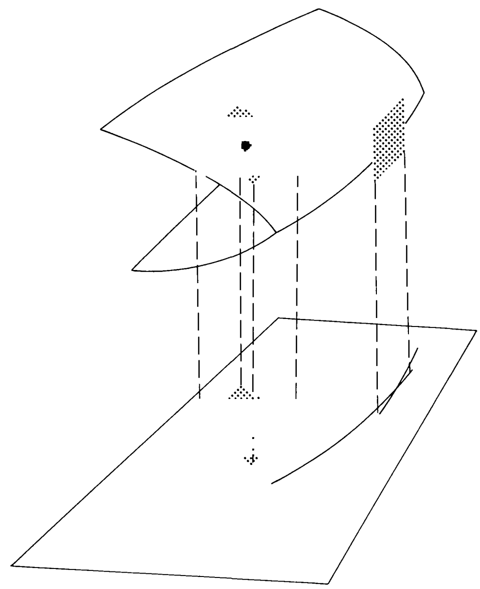


FIG. 1

**1. Introduction.** An easy way to visualise some types of differentiable mappings from a plane to a plane is to imagine the first plane being shaped into a smooth curved surface which is then projected downwards onto the other plane as in Fig. 1. As the mapping $f$ is differentiable, at each point $x$ it has a derivative $Df(x)$ which will be the $2 \times 2$ matrix of partial derivatives, providing a "linear approximation" to the mapping at each point. This approximation is seen by looking at what the derivative does to the tangent plane at the point $p$ on the surface. If we imagine the tangent plane at $p$ to be made of some opaque material and that a light is shining down from

*Jonathan Britt*: I am 36 years old and for the last twelve of those have been teaching mathematics at Southampton College of Higher Education. I was awarded my B.A. and M.Sc. at Watwick University in 1969 and 1971, respectively, and my Ph.D. at Southampton University in 1982. My research, which was carried on part time, was into the application of differential topology, in particular singularity theory, to mechanical systems with symmetry, such as the motion of a spinning top. I am still actively working in this field.

About a third of my teaching is Data Processing and Computing and I am particularly interested in the teaching of programming.

In my spare time I am involved both with Hospital Radio and an Amateur Dramatic group in the New Forest area.

My favourite way of relaxing is either to go for a ride in the New Forest (on a horse) or to go for a lesson to learn how to ride correctly! I have three children and an Afghan hound which I can occasionally be seen chasing, in a vain attempt to get him to come when I call!

above, then the image of a rectangular piece of the tangent plane can be visualised as the area in shadow on the plane below.

However, as in the example shown in Fig. 1, there may be points at which the tangent plane is vertical, and so the images of those tangent planes will be lines, not planes. (Analytically, these points are where the derivative has rank less than 2.) Such points are called *singular points* while points such as $p$, where the derivative maps tangent planes to planes, are called *regular points*. These terms will be defined precisely a little later. The mapping behaves very differently at singular points from the way it does at regular ones, where locally it behaves as if it were a mapping onto its codomain. Thus, in order to visualise fully the behaviour of a mapping, we need to understand what is happening at its singular points. This information provides a kind of skeleton on which to hang the clothing of regular points!

Sometimes when a mapping is perturbed slightly its singular points disappear entirely or are changed to another form, and in such cases the mapping is called *unstable*. If, however, the singular points cannot be removed or altered by perturbations of the mapping or, more precisely, if they are merely replaced by equivalent ones nearby, then the mapping is *stable*.

The natural question arises, can any unstable mapping be perturbed into a stable one? For mappings of the plane to itself the answer is yes, and in fact H. Whitney proved in his paper [**10**] that for differentiable mappings from $\mathbb{R}^n$ to $\mathbb{R}^p$, provided $n$ and $p$ are less than or equal to 5, the answer is yes as well.

When $n$ and/or $p$ are greater than 5 the position is somewhat more complicated and the detailed results can be found in J. Callahan's paper [**2**] which is a very easy to read and intuitive introduction to the study of singularities. For a more formal and detailed look at the theory see either the book by C. Gibson [**4**] or that by M. Golubitsky and V. Guillemin [**5**].

When studying a mapping, useful information can be gained by examining not only the set of singular points (the singularity set) itself but also the image of that set under the mapping.

For example, consider the mapping from the plane to itself given by:

$$f(x, y) = (x, xy - y^3).$$

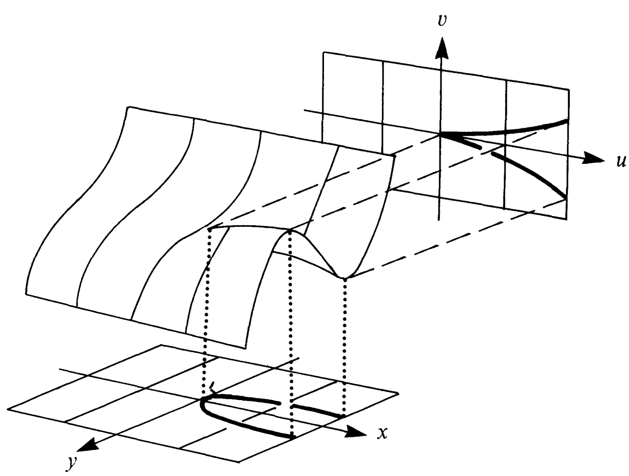



FIG. 2

A way of seeing this map is given in Fig. 2. The $(x, y)$-plane is first pleated at the origin and then projected onto the $(u, v)$-plane. Since the derivative of $f$ is

$$Df(x, y) = \begin{pmatrix} 1 & 0 \\ y & x - 3y^2 \end{pmatrix},$$

the singularity set, that is where the derivative has rank less than 2, is the parabola $x = 3y^2$. This

is the projection "downwards" of the fold lines in Fig. 2, together with the origin which corresponds to the pleat itself. The image of the singularity set is a curve with a cusp point whose equation is $4u^3 = 27v^2$, and it is this that has given the mapping the name of the *Whitney Cusp*.

Now let us look at the full inverse image of the cusp curve, that is all the points taken by $f$ to points on the cusp curve. This will help us to see more explicitly how $f$ behaves geometrically. In algebraic terms it will provide insight into the configurations of solutions of the equations

$$f(x, y) = (a, b)$$

for various choices of $a$ and $b$, because the number of solutions to those equations changes precisely where $(x, y)$ crosses the inverse image of the cusp curve. Note that this inverse image includes the singularity set with some other points besides.

To see those other points imagine parallel rays of light shining horizontally from all of the $(u, v)$-plane except the inside of the cusp curve. If the pleated surface is transparent, then a portion of that surface will be in darkness. Looking up from the $(x, y)$-plane to the surface, we see an area in shadow on the plane. The boundary of that area is the other part of the inverse image of the cusp curve. This is shown in Fig. 3.

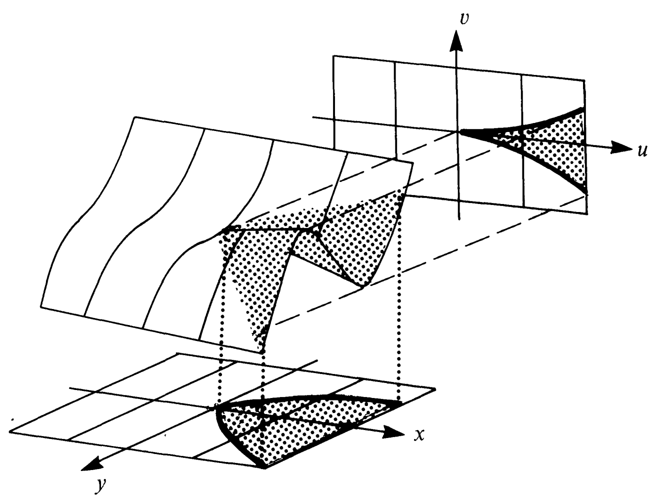



FIG. 3

Our interest in the Whitney cusp mapping arose, in fact, from studying the motion of a symmetric top spinning under gravity. The top can perform several different kinds of peculiar motion depending on its energy $E$ and angular momentum $J$ about the vertical. These two can be expressed in terms of four variables, namely the angle between the top's symmetry axis and the vertical axis, the time derivative of that angle and the angular velocities around those two axes. In this way we get a mapping from $\mathbb{R}^4$ to the $(E, J)$-plane, which close to its singularity set turns out to be equivalent to a generalisation of the Whitney cusp mapping that we looked at above. So information about the inverse image of the generalised cusp map gives us insight into how different values of the variables will affect the motion of the top. Further details can be found in [**1**].

Now we need some formal definitions and notation. We will consider differentiable mappings from $\mathbb{R}^n$ to $\mathbb{R}^p$, where $n \geqslant p$. Such a map $f$ has a derivative $Df(x)$ at each point $x$ of $\mathbb{R}^n$. The derivative is a $p \times n$ matrix whose $ij$th element is the partial derivative of the $i$th component of the mapping with respect to the $j$th variable. It necessarily has rank less than or equal to $p$. We classify singular points of such a map by the amount that the derivative loses rank at that point.

DEFINITION. The *singularity set* $S_k(f)$ is the set of points in $\mathbb{R}^n$ where the derivative of $f$ has rank exactly equal to $p - k$.

Next we will say what we mean by a stable mapping. To start with, we will call two maps from $\mathbb{R}^n$ to $\mathbb{R}^p$ *equivalent* when there are suitable coordinate changes in both the domain $\mathbb{R}^n$ and the codomain $\mathbb{R}^p$ which convert one map into the other. These coordinate changes will be differentiable maps with differentiable inverses; such maps are called diffeomorphisms. More precisely:

DEFINITION. Two maps $f$ and $g$ from $\mathbb{R}^n$ to $\mathbb{R}^p$ are *equivalent* if there are diffeomorphisms $h$ from $\mathbb{R}^n$ to itself, and $k$ from $\mathbb{R}^p$ to itself, such that $g = k \circ f \circ h$.

We are more often concerned with *local equivalence* where we require $g = k \circ f \circ h$ to hold only in the neighborhood of a particular point.

DEFINITION. A map $f$ from $\mathbb{R}^n$ to $\mathbb{R}^p$ is *stable* if every map $g$ sufficiently close to $f$ is equivalent to $f$. Local stability at a point is defined in the same way but using local equivalence.

To explain the term "sufficiently close" we need a topology on the set of differential maps from $\mathbb{R}^n$ to $\mathbb{R}^p$. Such topologies with reasonable properties exist, but it is beyond the scope of this article to go into the details of them. Suffice it to say here that for two maps to be considered "close" their values and their derivatives to some high order must all be close throughout the region under consideration. Full details can be found on page 42 of [**5**].

We are going to be looking exclusively (until Section 5) at mappings into the plane and so the only singularity sets that can exist are $S_1(f)$ and $S_2(f)$. However, it is implicit in the work done by Whitney in [**10**] that:

THEOREM. *If $f$ is a mapping from $\mathbb{R}^n$ to the plane and $S_2(f)$ is non-empty, then $f$ is unstable.*

In other words, if we have a singular point belonging to $S_2(f)$, then by perturbing the mapping we can either change the singularity to one belonging to $S_1(f)$ or destroy it altogether. So we only need to study singularities belonging to $S_1(f)$.

*Sketch of the proof of the Theorem.* Since $Df(x)$ is given as a $2 \times n$ matrix, we look at the space of all $2 \times n$ matrices. In that space the set of all matrices of rank 0 is just one point, the zero matrix, while the set of matrices with rank 1, which is given by $n - 1$ equations, is a "hypersurface" or manifold $T$ of codimension $n - 1$. We can think of $x \mapsto Df(x)$ as a mapping from $\mathbb{R}^n$ into the space of matrices, giving an $n$-dimensional manifold $M$ parametrised by $x$. Typically such a manifold $M$ will meet a codimension $n - 1$ manifold $T$ in a 1-dimensional curve but will not contain the zero matrix point at all. If $S_2(f)$ is not empty, then the manifold $M$ does contain the zero point and a small perturbation will, by the remark above, cause the manifold to miss that point and thus destroy the $S_2(f)$ singularity. In contrast, were $M$ meets the manifold $T$ of rank 1 matrices and does so in general position, then a small perturbation will change the point, but not the fact, of intersection. This shows that singularities belonging to $S_2(f)$ are unstable, so that stable mappings can have singularities belonging only to $S_1(f)$. ■

In the case of a mapping into the plane the set $S_1(f)$ will consist of curves and in the more general case of a mapping from $\mathbb{R}^n$ to $\mathbb{R}^p$ it will be a $p - 1$ dimensional manifold.

Now in order to scrutinize the singularities of $f$ more closely we restrict $f$ to $S_1(f)$, on which it has rank at most $p - 1$, and see at what points, if any, $f$ restricted to $S_1(f)$ drops rank by 1. (We are in fact using one form of the implicit function theorem which tells us that we may perform these operations.) The set of such points is denoted $S_{1,1}(f)$. This idea can be generalised; for details of how see pages 174–190 of [**4**].

After this groundwork we go back to our example and examine the cusp mapping in more detail.

**2. The Whitney cusp mapping from the plane to itself.** In 1955 H. Whitney [**9**] proved that the most complicated singularity that can arise in a stable mapping of the plane to itself is locally equivalent to the mapping which we have already met, that is:

$$f(x, y) = (x, xy - y^3).$$

From our previous discussion we can see that the set $S_1(f)$ is just the parabola $x = 3y^2$ and the set $S_{1,1}(f)$ is the origin itself. The image of $S_1(f)$ under $f$ is the cusp curve $4u^3 - 27v^2 = 0$ and the image of $S_{1,1}(f)$ is the actual cusp point $u = v = 0$.

Points inside the cusp curve are the image of three distinct points in the $(x, y)$-plane, while those outside have only one inverse image point. Those on the cusp curve, except for the origin, have two distinct inverse image points one of which lies on the singularity set and the other on the second curve we found in the Introduction.

Finding the equation of this second curve is straightforward. We replace $u$ by $x$ and $v$ by $xy - y^3$ in the equation of the cusp curve, to get

$$4x^3 - 27(xy - y^3)^2 = 0,$$

that is

$$(x - 3y^2)^2(4x - 3y^2) = 0.$$

Hence the equation of this second curve is $4x = 3y^2$ and it too is a parabola.

Thus the complete inverse image of the cusp curve consists of these two parabolae. We next investigate an analogous cusp mapping from $\mathbb{R}^3$ to the plane, and we shall find that these two parabolae embed in the new inverse image of the cusp curve in a most interesting and unexpected way.

**3. The cusp mapping from $\mathbb{R}^3$ to the plane.** Again for a stable mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$ the most complicated singularity that can exist is a so-called cusp point. In his paper of 1958 **[10]**, Whitney gave a standard form for this mapping, namely

$$f(x, y, z) = (x, xy - y^3 - z^2),$$

where now the cusp point is at the origin.

By means of a similar calculation as was done for the mapping from the plane to itself, it is easy to see that

$$S_1(f) = \{(x, y, 0);\ x = 3y^2\}$$

and

$$S_{1,1}(f) = \{(0,0,0)\}.$$

As before, the image of $S_1(f)$ is the cusp curve $4u^3 - 27v^2 = 0$. Since we are dealing with a mapping from $\mathbb{R}^3$ to the plane, the inverse image of a point will in general be a curve. This will be true for points on the cusp curve itself as well, although we might expect the curves we get from such points to be "less well behaved" in some sense than those that come from points not on that curve.

We will start by finding the inverse image of the origin. A simple calculation shows that this is given by

$$\{(0, y, z);\ y^3 = -z^2\},$$

which is a cusp curve lying in the $(y, z)$-plane.

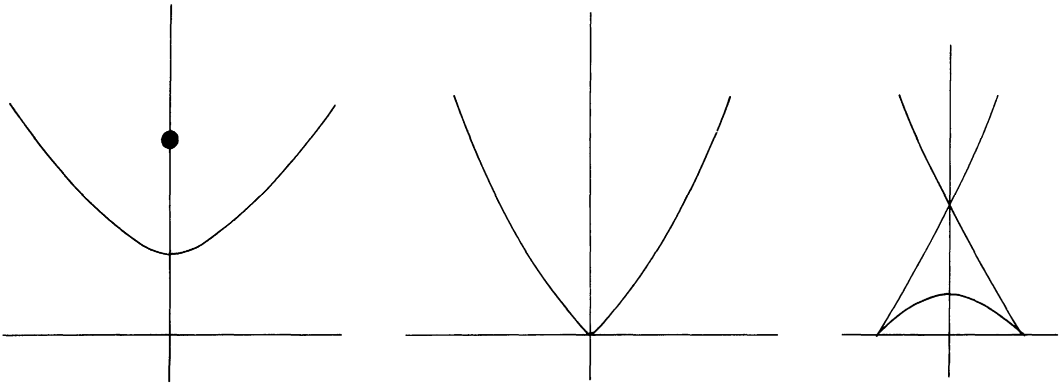
If we were to restrict the mapping to the plane $z = 0$ we would be back in the situation we discussed in Section 2. Hence there will be two parabolae lying in that plane and mapping onto the cusp curve, namely $S_1(f)$ and

$$\{(x, y, 0);\ 4x = 3y^2\}.$$

We can write down the equation of the full inverse image of the cusp curve, just as we did in Section 2. It is:

$$4x^3 - 27(xy - y^3 - z^2)^2 = 0. \tag{3.1}$$

To get some idea of what this might look like it is helpful to take slices parallel to the

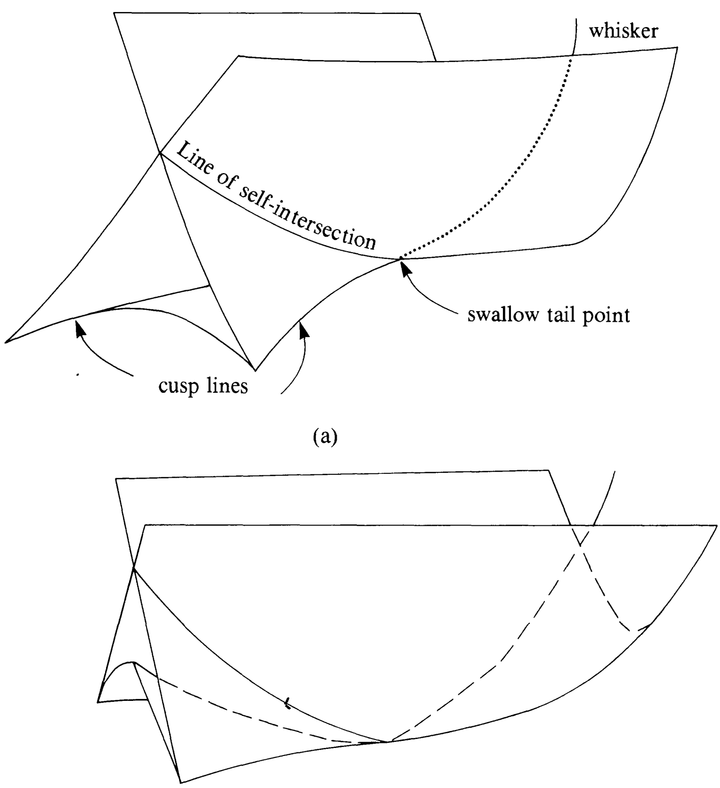


$y > 0$
(A double isolated point exists.)

$y = 0$

$y < 0$

FIG. 4




(a)

(b)

FIG. 5

$(x, z)$-plane. Choosing different constant values of $y$, we get pictures of the kind shown in Fig. 4. These look like cross sections of the swallowtail as shown in Callahan [**3**] or on page 177 of Poston and Stewart [**8**] and as sketched in Fig. 5(a). In the two references given above the swallowtail is described as the bifurcation set of the real swallowtail catastrophe. However it is a classical object extensively studied in the last century as a set of points in $(a, b, c)$ space for which the quartic

equation

$$x^4 + ax^2 + bx + c = 0$$

had repeated roots, real or complex.

There is a slight difference between the two uses—in the second case there is an isolated whisker (corresponding to repeated complex roots) which together with the line of self intersection forms a parabola. In the bifurcation set picture this whisker does not appear, because these complex roots are irrelevant to real bifurcations. For a fuller discussion of the whisker we refer the reader to pages 130–132 of Poston and Stewart [7]. (The whisker will in fact turn out to be half of $S_1(f)$, the other half being the line of self intersection when $y < 0$.)

The double isolated point which occurs in Fig. 4 when $y > 0$ corresponds to this whisker, so we suspect that the inverse image of the cusp curve as given by Equation 3.1 will be equivalent to the set of points in $(a, b, c)$ space where

$$x^4 + ax^2 + bx + c = 0$$

has repeated roots. This set is given algebraically by the vanishing of the discriminant $\Delta$, where

$$\begin{aligned} \Delta &= S^3 - 27T^2, \\ S &= \frac{a^2}{12} + c, \qquad (3.2) \\ T &= \frac{ac}{6} - \frac{b^2}{16} - \frac{a^3}{216}. \end{aligned}$$

See for example page 120 of [7]. In fact we can prove:

PROPOSITION. *There is an origin preserving diffeomorphism from* $\mathbb{R}^3$ *to itself which takes the set given by Equation* 3.1 *to that defined by Equation* 3.2.

*Proof.* A comparison of the two equations shows us that we may define the mapping by the bijection:

$$\begin{aligned} x &= 4^{-1/3}\left(\frac{a^2}{12} + c\right), \\ y &= 4^{1/3}\frac{a}{6}, \\ z &= \frac{b}{4}. \end{aligned}$$

A quick calculation shows that this map has non-vanishing Jacobian, hence is a diffeomorphism, and that it converts Equation 3.1 to Equation 3.2 and thus the proposition is proved. ∎

We now know that the inverse image of the cusp curve is a swallowtail like that shown in Fig. 5(a) but bent upwards slightly to the shape shown in Fig. 5(b). However, we would like more detail as to which points of the inverse image correspond to individual points on the cusp curve. To get this, using the fact that under the mapping $u$ is just equal to $x$, we intersect the cusp curve with a line $u = k^2$. The inverse image of this line is just the plane $x = k^2$, and so the curves in which that plane intersects the swallowtail will be the inverse images of the two points on the cusp curve and the line $u = k^2$. Fig. 6(a) shows the intersection of the plane and the swallowtail consisting of two curves and a double isolated point.

The coordinates of the two points where the line $u = k^2$ meets the cusp curve are $(k^2, 2k^3/3\sqrt{3})$ and $(k^2, -2k^3/3\sqrt{3})$, (where $k^3$ is taken to be positive). So the inverse images of the two points are respectively

$$z^2 = -\left(y - \frac{k}{\sqrt{3}}\right)^2\left(y + \frac{2k}{\sqrt{3}}\right)$$

and

$$z^2 = -\left(y + \frac{k}{\sqrt{3}}\right)^2\left(y - \frac{2k}{\sqrt{3}}\right).$$

These curves are sketched in Fig. 7, and it is easy to see that the inverse image of the point $b$ with coordinates $(k^2, -2k^3/3\sqrt{3})$ is the curve with a point of self intersection whilst that of the point $a$ with coordinates $(k^2, 2k^3/3\sqrt{3})$ is the other curve together with the double isolated point.

Having found the inverse image of points on the cusp curve we turn our attention to points lying inside and outside of that curve. Again we take the line $u = k^2$ and look at points on the line. Their inverse images will be in the plane $x = k^2$ as before. Consideration of Figs. 6(a) and 7, and a little bit of simple algebra, enables sketches of the inverse images to be made easily. These, together with the inverse images of points on the cusp curve, are shown in Fig. 8. Points outside the cusp curve will have only one curve as their inverse image while points inside will have two.

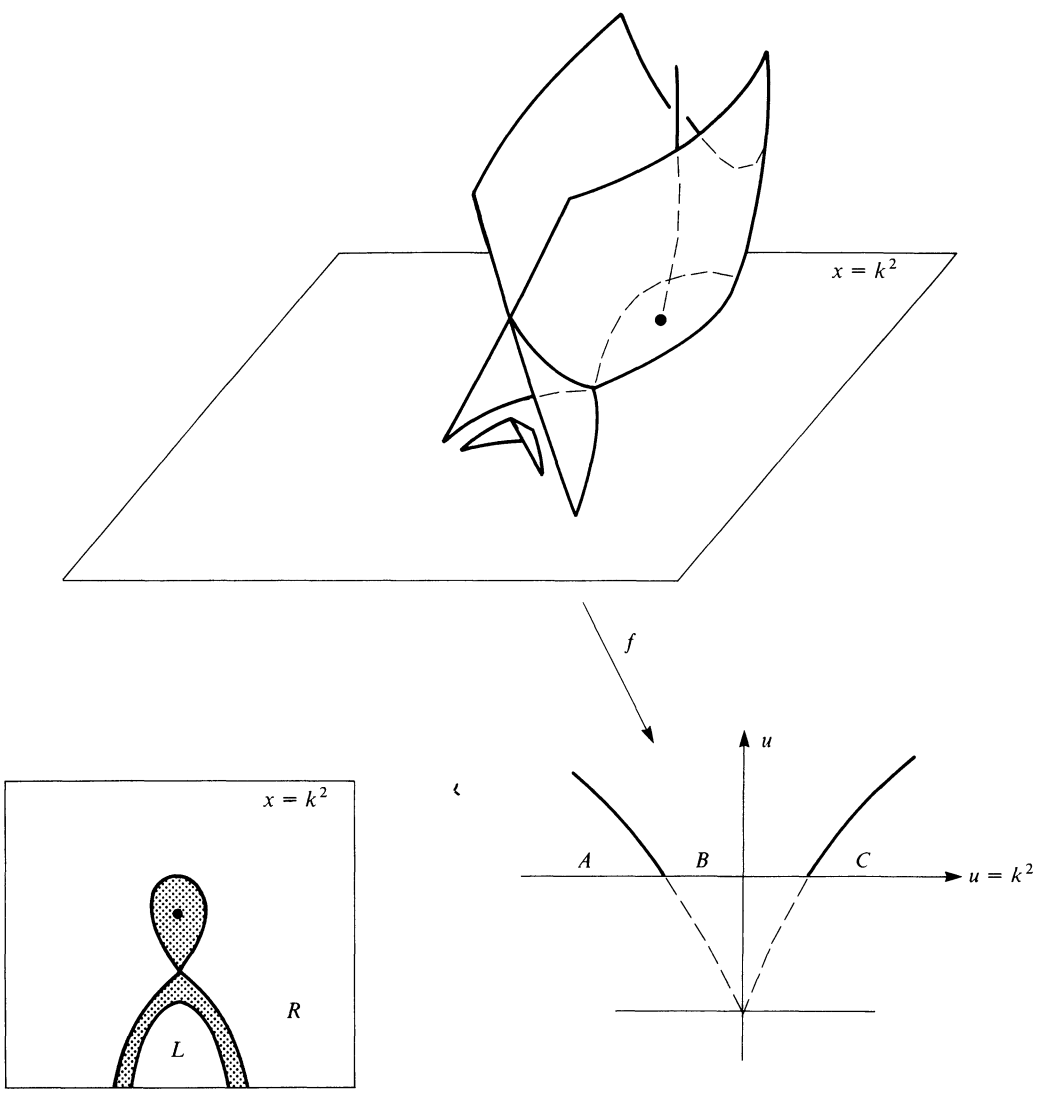



FIG. 6(a)

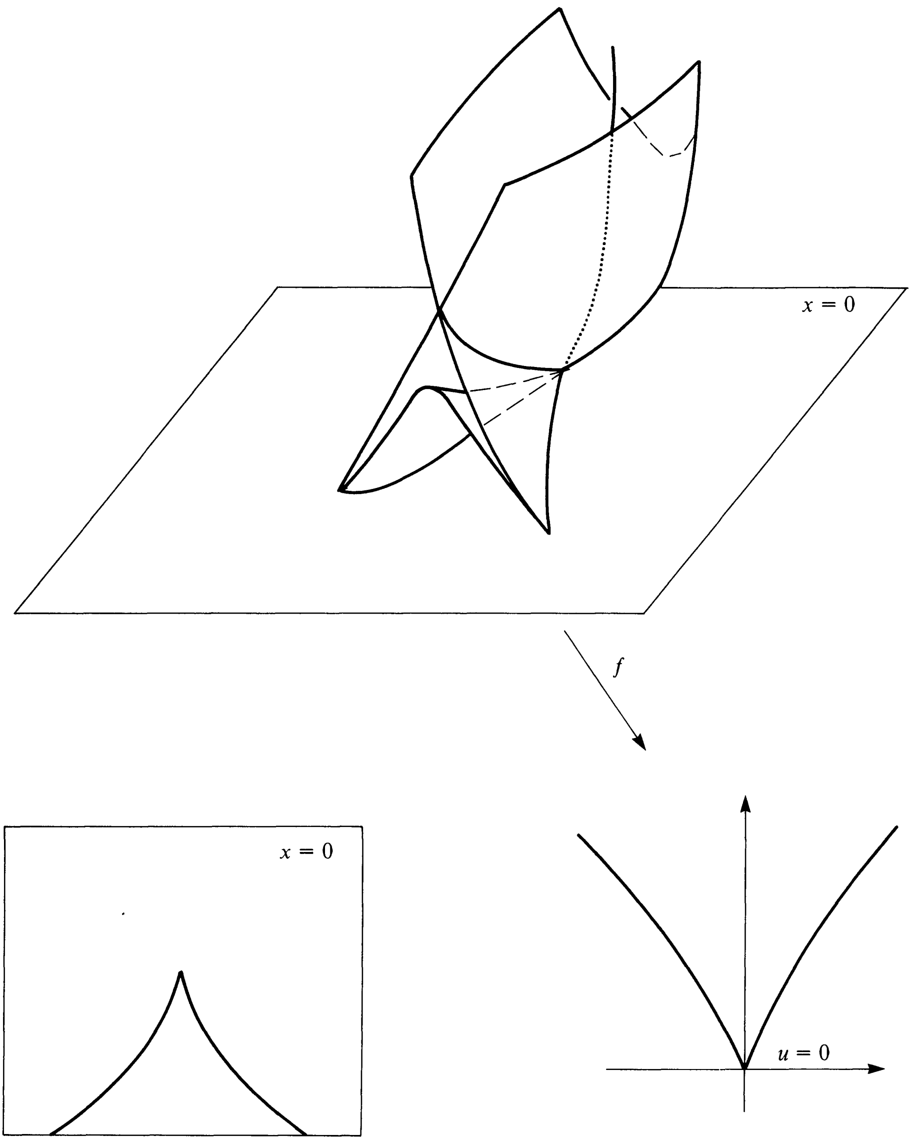



FIG. 6(b)

One of those will be a closed curve around the whisker and the other an open curve in the "tail portion" of the space enclosed by the swallowtail.

Restricting to $z = 0$, the behaviour described in Section 2 of points outside having an inverse image of one point and points inside having three points in their inverse image can be clearly seen. Furthermore, Fig. 8 shows that the inverse images of points not on the cusp curve are regular curves, but for points on the cusp curve the inverse images are curves with singular points, either a double isolated point or a point of self-intersection. (Again we are using the implicit function theorem here.) This illustrates our earlier remark about expecting the inverse image of the cusp curve points to be "less well behaved" in some way.

We can summarise all of this by looking back at Fig. 6 and imagining a tide rising around the swallowtail. Before the tide reaches the swallowtail, that is when $x = -k^2$, the intersection of the plane and the swallowtail is empty as is that of the line $u = -k^2$ and the cusp curve. This is

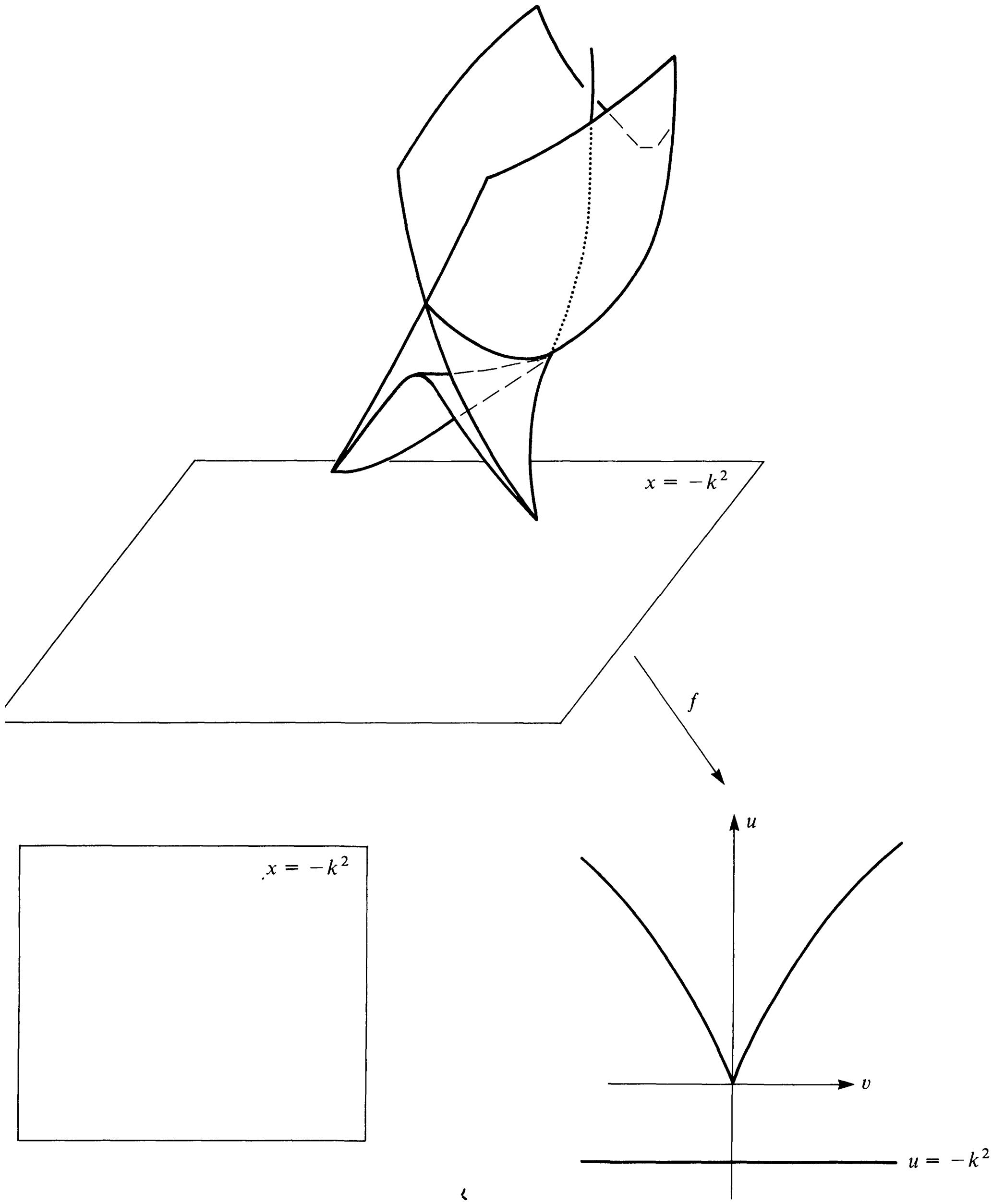



FIG. 6(c)

shown in Fig. 6(c). The tide rises until the water just starts to lap at the feet of the swallowtail, then the cusp curve on which the swallowtail sits is clearly seen as the inverse image of the cusp point in the $(u, v)$-plane, see Fig. 6(b). The inverse images of the two half lines $v > 0$ and $v < 0$ are the two regions of the plane $x = 0$ separated by the cusp curve.

As the tide rises further we get the picture shown in Fig. 6(a). The two points where the cusp curve intersects the line $u = k^2$ have the two curves with the isolated point that we described above as their inverse image. These two points divide the line $u = k^2$ into 3 segments labelled $A$, $B$ and $C$ in Fig. 6(a). The inverse image of the line segment $A$ is the region of the plane labelled $L$

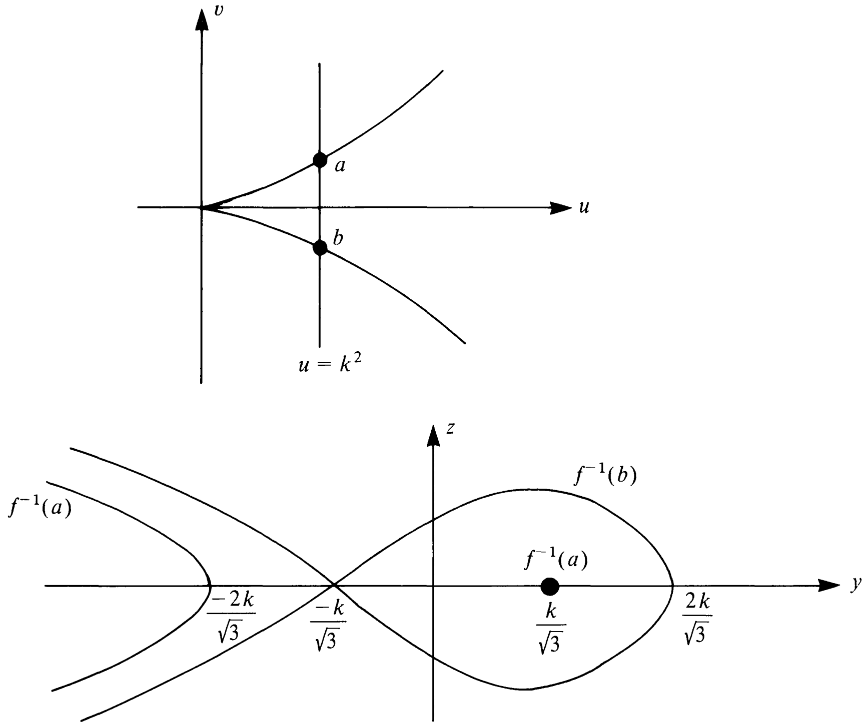



FIG. 7

while that of the line segment $C$ is the region $R$. The two regions shown shaded form the inverse image of the central Section $B$ of the line.

Finally notice how as the rising tide changes Fig. 6(b) to Fig. 6(a) the cusp point in the plane $x = 0$ bifurcates into four separate points, two regular and two singular, while the cusp curve becomes two separate curves.

**4. The cusp mapping from $\mathbb{R}^n$ to the plane.** The general form for the Whitney cusp mapping from $\mathbb{R}^n$ to the plane, as given in [**10**], is

$$f(x, y, z_1, \ldots, z_{n-2}) = \left(x, xy - y^3 - z_1^2 \pm z_2^2 \pm \cdots \pm z_{n-2}^2\right).$$

In fact in Whitney's paper there is a "$\pm$" sign in front of $z_1^2$ as well, but we choose the minus sign to maintain the convention used in Section 3. There the two possible choices give an equivalent geometric picture of the inverse image (by reflection in the $xz$ plane and the $u$ axis). In this section what is important is whether the signs in front of the $z^2$ terms are all the same.

If all the signs are the same then the cusp mapping can be put in the form:

$$f(x, y, z_1, \ldots, z_{n-2}) = \left(x, xy - y^3 - z_1^2 - \cdots - z_{n-2}^2\right).$$

We refer back to Fig. 8. The inverse image of the point 4 on the cusp curve will consist of an isolated point $(k^2, k/\sqrt{3}, 0, \ldots, 0)$ and a "manifold of revolution" formed by revolving the curve shown in Fig. 8(d) as the inverse image of the point 4 around the $y$-axis in the hyperplane $x = k^2$. Similarly the inverse image of the points 1, 2, 3 and 5 in Fig. 8 will be obtained by revolving around the $y$-axis their inverse images as given in Fig. 8, again in the hyperplane $x = k^2$. Hence no essentially new features appear in the geometry of the inverse image set in this case.

However, if some of the signs in front of the $z^2$ terms are plus and some minus then, by renumbering the coordinates if necessary, we can write the mapping as

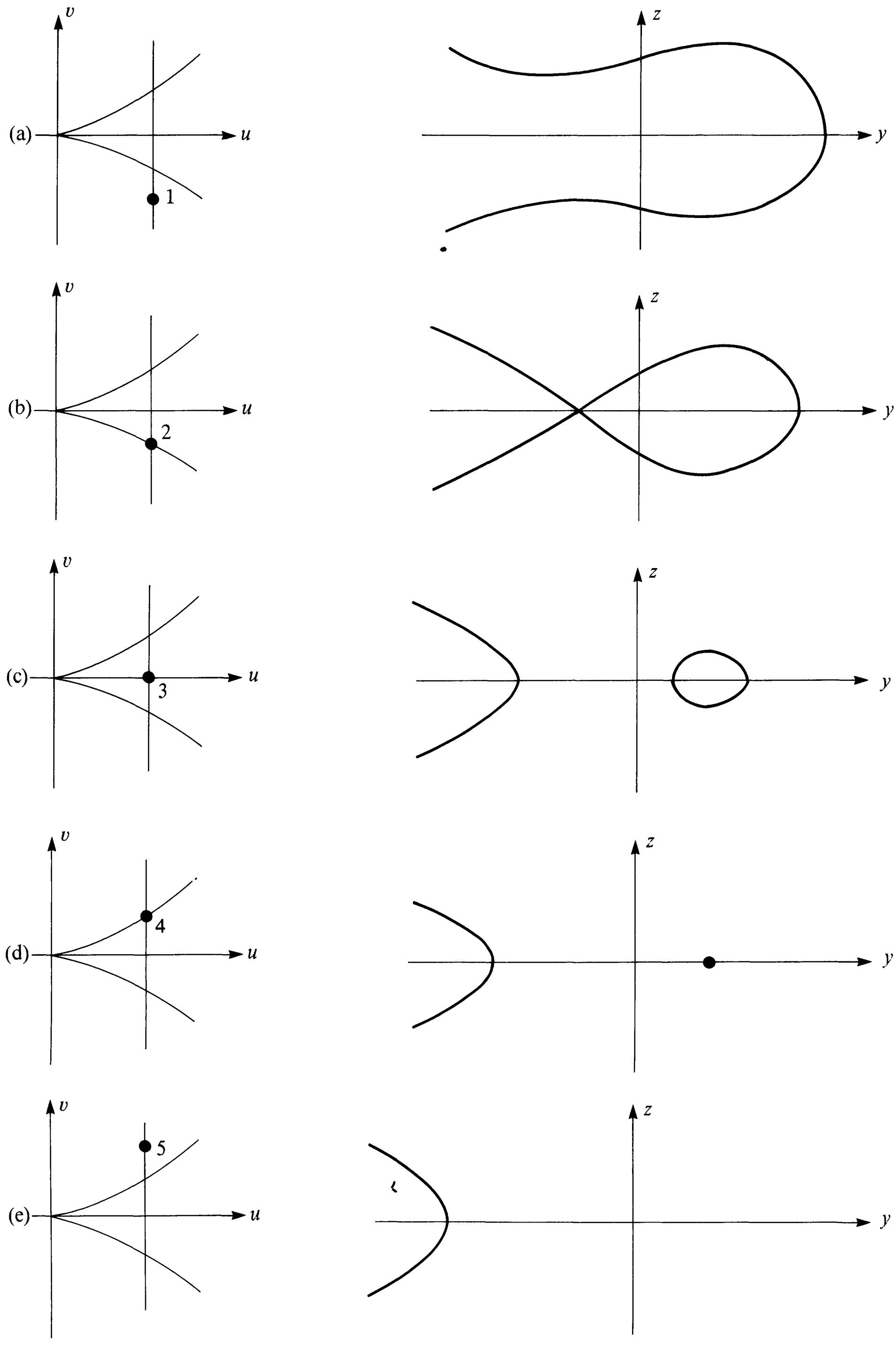



FIG. 8

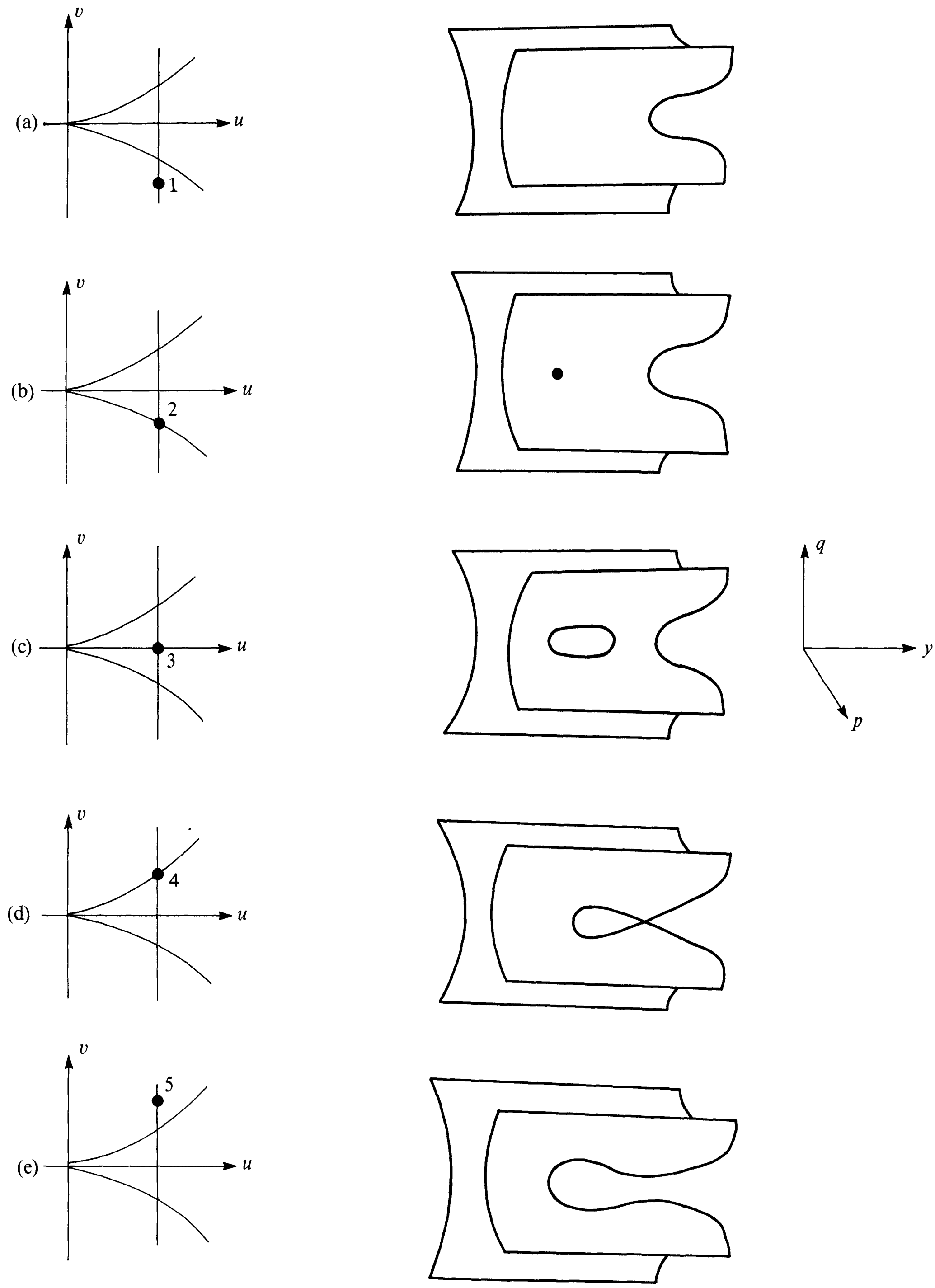



FIG. 9

$$f(x,y,z_1,\dots,z_{n-2}) = \left(x, xy - y^3 - z_1^2 - \cdots - z_r^2 + z_{r+1}^2 + \cdots + z_{n-2}^2\right),$$

or as

$$f(x,y,z_1,\dots,z_{n-2}) = \left(x, xy - y^3 - p^2 + q^2\right),$$

where

$$p^2 = z_1^2 + \cdots + z_r^2,$$

and

$$q^2 = z_{r+1}^2 + \cdots + z_{n-2}^2.$$

Hence we can regard the mapping as being the composition of a cusp mapping $g$ from $\mathbb{R}^4$ to the plane and a mapping $h$ from $\mathbb{R}^n$ to $\mathbb{R}^4$ given by

$$g(x,y,p,q) = \left(x, xy - y^3 - p^2 + q^2\right)$$

and

$$h(x,y,z_1,\dots,z_{n-2}) = (x,y,p,q).$$

The inverse image under $h$ of any point in the subset $K$ of $\mathbb{R}^4$, where $K$ is given by:

$$K = \{(x,y,p,q); p,q \neq 0\},$$

is just the "torus" $S^{r-1} \times S^{n-r-3}$. However, for a point in $\mathbb{R}^4$ with either $p$ or $q = 0$, the inverse image degenerates into a sphere, $S^{n-r-3}$ or $S^{r-1}$, respectively. Hence, once we have found the inverse image of a point under the cusp mapping $g$, we take the Cartesian product of points that lie in the intersection of the inverse image and $K$ with the "torus," but for those points not in $K$ we take the product with the appropriate sphere to generate the inverse image under $f$. (Of course if a point in $\mathbb{R}^4$ has $p$ and $q = 0$, then its inverse image under $h$ will just be one point.) We will refer to this process as forming a "generalised manifold of revolution".

It is straightforward to show that the inverse images under the mapping $g$ of the points 1, 2, 3, 4 and 5 on the line $u = k^2$ are as shown in Figs. 9(a) to (e).

These pictures have several interesting features. Firstly, inverse images of a point are unbounded in the positive $y$ direction in contrast to the position in Section 3, where the inverse images always had $y$ less than or equal to a particular value that depended on that point. Secondly, notice that the inverse images of all points (including the points 3 and 4) are now connected sets.

We can see too that there is a duality in the pictures in the following sense: if we take the inverse image of the point 1 and rotate it through $90°$ in the $(p,q)$ plane and then through $180°$ in the $(y,p)$ plane, we arrive at the inverse image of the point 5. The inverse images of the points 2 and 4 are dual in the same way, while that of 3 is self dual. These geometric facts of course correspond to the equivalence between replacing $y$ by $-y$ and interchanging $p$ and $q$ in $\mathbb{R}^4$ and replacing $v$ by $-v$ in $\mathbb{R}^2$ under the mapping $g$.

Lastly, let us see what happens if we intersect each of the sets with planes $q =$ constant. Because of the symmetry of the pictures we can restrict our attention to $q \geqslant 0$.

For the set in Fig. 9(a), when $q = 0$ we have a curve like that shown in Fig. 8(a). As $q$ increases, the curve fills out its "waist" and expands, but remains a simple open curve. In part (b) of Fig. 9, for $q = 0$ we have the curve shown in Fig. 8(b), but for $q > 0$ we obtain a curve like that shown in part (a) of that figure. In the case of Fig. 9(c), for $q = 0$ we have the two curves shown in Fig. 8(c). As $q$ increases the open and closed curves approach each other until they collide to form the curve in Fig. 8(b), then as $q$ continues to increase we again get curves like those seen in Fig. 9(a). This pattern holds true for all the pictures until finally for the inverse image of point 5 as $q$ increases from 0 all the different curves shown in Fig. 8 appear in order. (Of course the two curves that are the inverse images of points 2 and 4 appear only instantaneously.) Fig. 10 illustrates these curves appearing by showing a tide rising around the surface of Fig. 9(e).

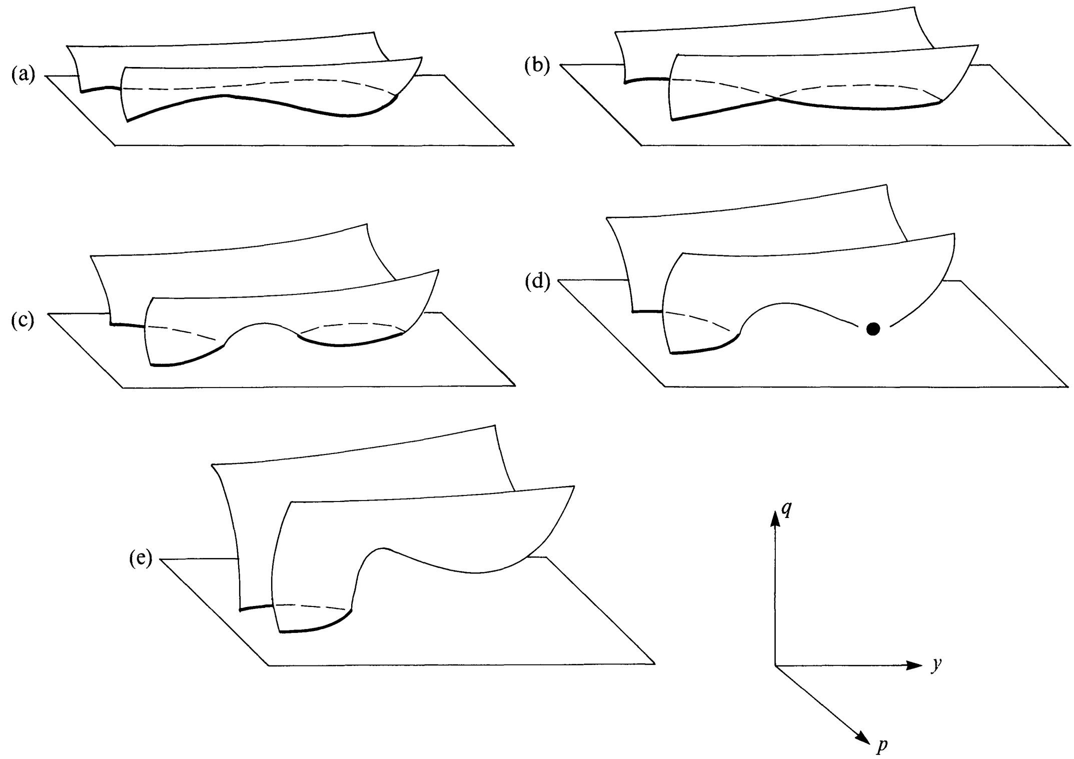



FIG. 10

To see algebraically why this happens, compare the map $g$ with the cusp map from $\mathbb{R}^3$ to the plane (Section 3) and take various different values of $q^2$.

In what we have discussed above we have covered the main features of the real geometry that can occur in the inverse image of a Whitney cusp mapping from $\mathbb{R}^n$ to the plane. The inverse image of the cusp curve will consist of either a "manifold of revolution" of one of the curves in Fig. 8(b) or (d) or a "generalised manifold of revolution" of one of the surfaces in Fig. 9(b) or (d).

**5. Further directions.** The cusp is the most complicated stable singularity that occurs in mapping from $\mathbb{R}^n$ to the plane. However, if we look at maps from $\mathbb{R}^3$ to itself another more complicated singularity can occur, namely the *swallowtail.* This is given up to equivalence by

$$f(x, y, z) = (x, y, xz + yz^2 + z^4).$$

Staying with the equidimensional case, as the dimension increases by one there is precisely one new type of stable singularity which appears. If we now increase the dimension of the domain but leave that of the codomain fixed, then the number of different types of singularities remains the same. An explicit formula for a general form for all these mappings can be found in the paper by B. Morin [**6**].

A possible next step then would be to examine the inverse images of the singularity sets of these higher dimensional singularities. For the swallowtail mapping from $\mathbb{R}^3$ to itself the approach that we have used so far is quite feasible, but once we go beyond that case the algebra involved becomes fairly forbidding and a more general approach might offer more chance of success.

So we will conclude by giving a very brief description of what does happen in the case of the swallowtail mapping from $\mathbb{R}^3$ to itself. With $f$ as given above, the singularity set $S_1(f)$ is the cubic surface with equation

$$4z^3 + 2yz + x = 0$$

shown in Fig. 11(a), and its image under $f$ is the swallowtail set we saw in Fig. 5(a) but without the whisker. (This set is redrawn in Fig. 11(b).)

When it comes to finding the inverse image of this singularity set, the easiest way to proceed is to take $v$ = constant slices and look at the inverse images of the curves that this gives. For $v \geqslant 0$ no new points at all appear.

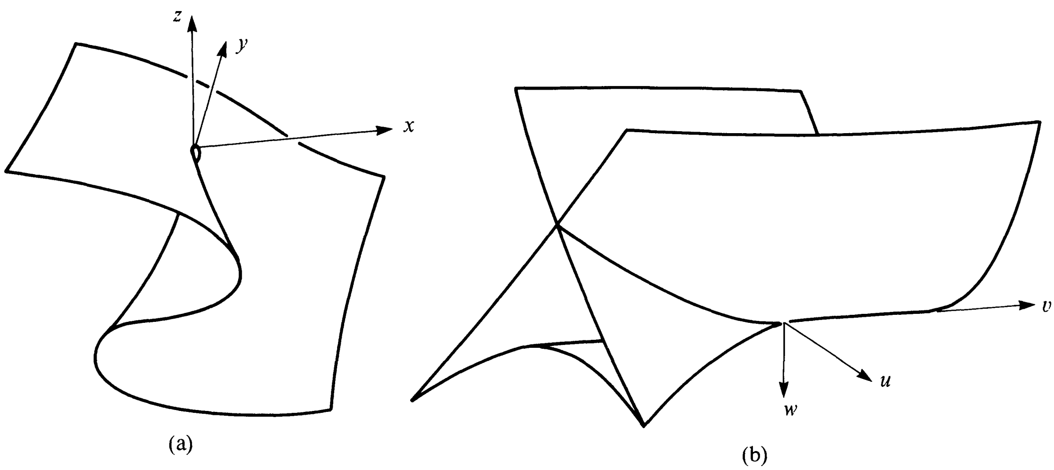



FIG. 11

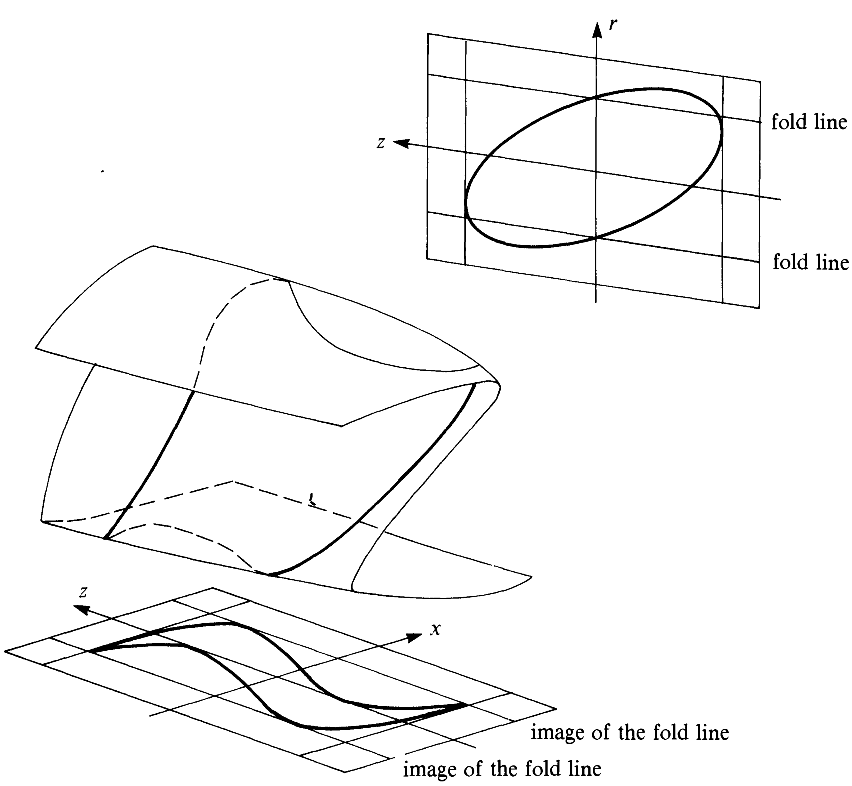



FIG. 12. (The reader is encouraged to take a piece of acetate and draw on it the ellipse. Then by experimenting with folding the sheet and looking down from above he will see how the double duck billed loop is formed.)

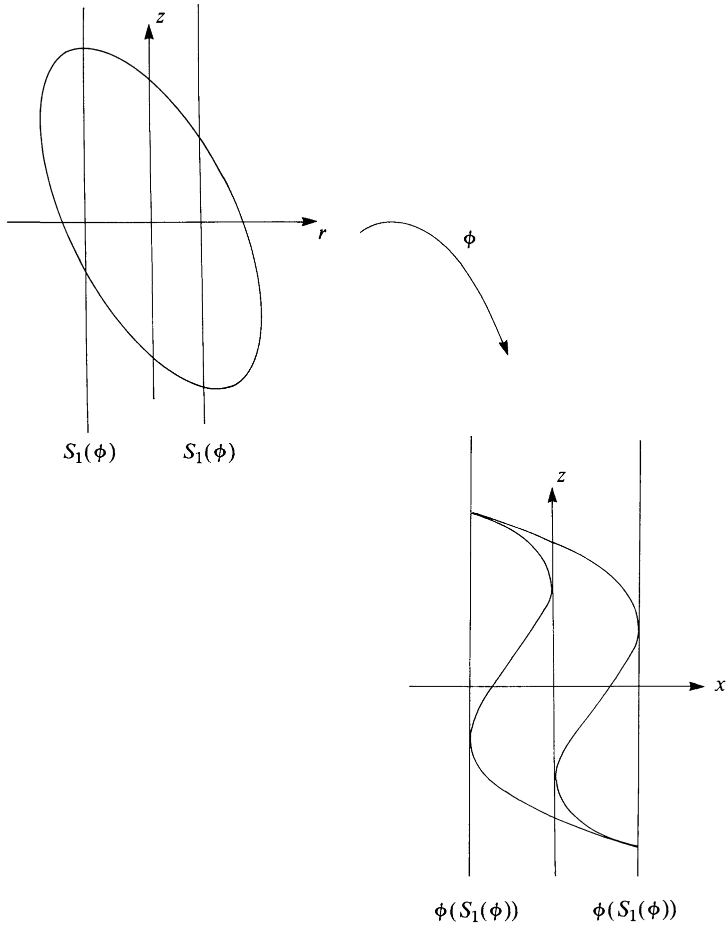



FIG. 13

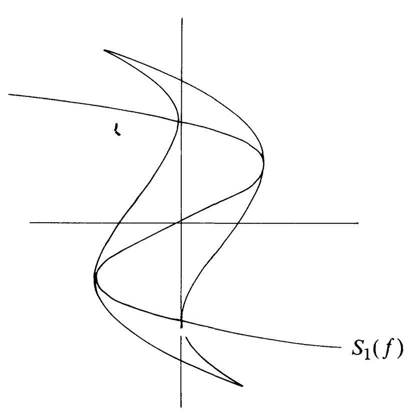



FIG. 14

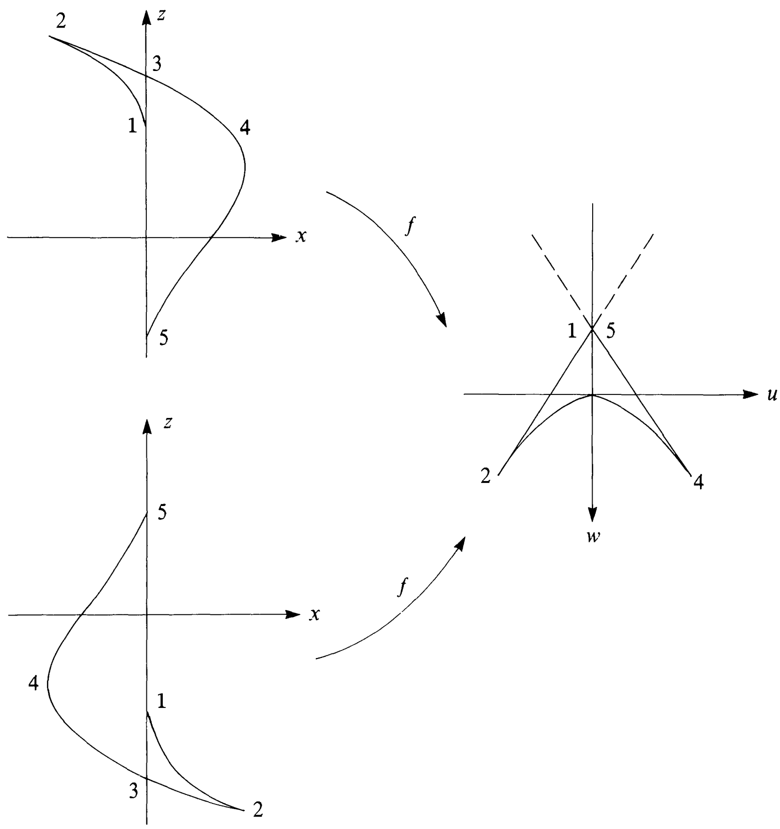



FIG. 15

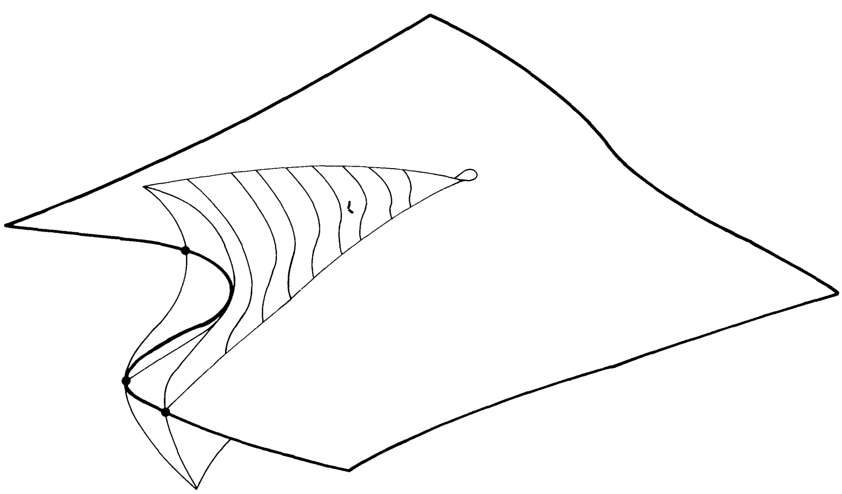


FIG. 16

However, if we take a slice $v = -k^2/2$ through the swallowtail and try to find the inverse image of the curve that this gives, we find that those points not belonging to $S_1(f)$ are given by the equations:

$$x = k^2 r - 4r^3$$

and

$$z = -r \pm \sqrt{\frac{k^2 - 4r^2}{2}}\,.$$

Although these equations seem at first difficult to interpret geometrically, we can do so by rewriting the second equation as

$$z^2 + 2rz + 3r^2 = \frac{k^2}{2},$$

which gives an ellipse in the $(r, z)$ plane.

Then if we consider a mapping $\phi$ from the plane to itself defined by

$$\phi(r, z) = (k^2 r - 4r^3, z),$$

the set of points we are interested in is the image of the ellipse under $\phi$. If we analyse $\phi$ using the techniques discussed above, we find that $\phi$ has two lines of fold points, which together make up $S_1(\phi)$, and no other singularities at all. The set $S_1(\phi)$ is given by

$$S_1(\phi) = \left\{(r, z);\ r = \pm\frac{k}{2\sqrt{3}}\right\}.$$

Fig. 12 gives a visual interpretation of how the mapping $\phi$ behaves. In particular, notice that when the ellipse crosses the fold line obliquely the image of the ellipse is a line tangential to the image of the fold line, but where the ellipse crosses perpendicular to the fold line the image of the ellipse is a cusp (though not, however, a cusp arising from a Whitney cusp map!) Fig. 13 gives an accurate drawing of both the ellipse and its image under $\phi$, a double duck billed loop.

The complete inverse image of the $v = -k^2/2$ slice through the swallowtail is drawn in Fig. 14. For details of how the duck billed loop maps to the swallowtail see Fig. 15. Putting this information together enables us to sketch the entire inverse image of the swallowtail set in Fig. 16.

**Acknowledgements and Dedication.** I would like to thank David Chillingworth for all the help and advice he gave to me in the writing of this paper, which is dedicated to the memory of my father Albert Britt who died suddenly on 27th January 1984.

## References

**1.** J. P. Britt, Singularity theory and geometry in the motion of a top, Ph.D. Thesis, Southampton University, 1982.

**2.** J. J. Callahan, Singularities and plane maps, this MONTHLY, 81 (1974) 211–240.

**3.** ______, Singularities and plane maps II, this MONTHLY, 84 (1977) 765–803.

**4.** C. G. Gibson, Singular Points of Smooth Mappings, Research notes in Mathematics 25, Pitman, London, 1979.

**5.** M. Golubitsky and V. Guillemin, Stable Mappings and Their Singularities, Graduate Texts in Mathematics 14, Springer-Verlag, New York, 1973.

**6.** B. Morin, Formes canoniques des singularités d'une application différentiable, C. R. Acad. Sci. Paris, t260 (1965) 5662–5665, 6503–6506.

**7.** T. Poston and I. N. Stewart, Taylor Expansion and Catastrophes, Research Notes in Mathematics 7, Pitman, London, 1976.

**8.** ______, Catastrophe Theory and its Applications, Pitman, London, 1978.

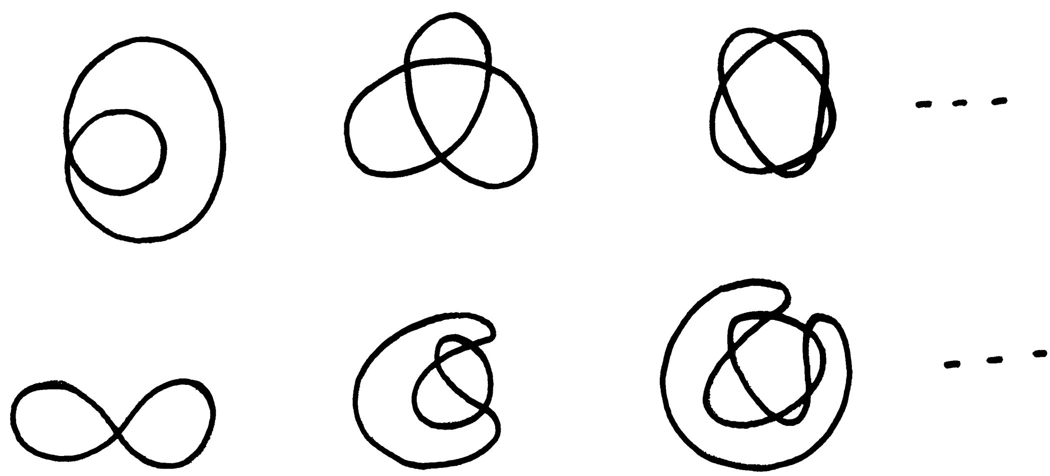
**9.** H. Whitney, On Singularities of mappings of Euclidean spaces, I. Mappings of the plane into the plane, Ann of Math., 62(1955) 374–410.

**10.** ______, Singularities of mappings of Euclidean spaces, in Symposium internacional de topologia algebraica, 285–301, Universidad Nacional Autonoma de Mexico and UNESCO, Mexico City, 1958.

# CHARACTERIZING MAXIMALLY LOOPED CLOSED CURVES

F. J. CRAVEIRO DE CARVALHO*
*Departamento de Matemática, Universidade de Coimbra, 3000 Coimbra, Portugal*

**1. Introduction.** It is easy to convince oneself after a few minutes of experimental sketching that there are no analogous curves to those shown in Fig. 1 with an *even* number of double points.



$n = 1$ $\qquad$ $n = 3$ $\qquad$ $n = 5$

FIG. 1

We shall prove that this is indeed the case as a corollary to Theorem 3.1. The relevant property of the curves of Fig. 1 is that they are "maximally looped". We also calculate the rotation number of maximally looped curves, and compute the number of double tangents for those without inflection points.

**2. Definitions.** A *closed curve* is a $C^1$ immersion $f\colon S^1 \to R^2$. A pair $\{x, y\}$ of distinct points of $S^1$ is a *double point*, *crossing*, or *self-intersection* of $f$ if $f(x) = f(y)$. We assume that no three distinct points of $S^1$ have the same image and that $f$ is *self-transverse*. By self-transverse we mean that if $\{x, y\}$ is a double point, then the tangents to $f$ at $x$ and $y$ are distinct. These restrictions are not very serious. In fact, with a suitable topology for the set of maps $g\colon S^1 \to R^2$, the immersions we are interested in form an open and dense subset. According to **[1]** and **[14]**, they are called *clean* immersions and *normal* curves respectively. In this note the immersions are always non-injective.

By composing with $g\colon R \to S^1$ given by $g(t) = (\cos 2\pi t, \sin 2\pi t)$, we can consider $f$ to be a periodic map $f\colon R \to R^2$ with period 1.

A *loop* of $f$ is a restriction of $f$ to a closed interval $[a, b]$ such that $f|[a, b)$ is injective and $f(a) = f(b)$. We shall not distinguish between two loops of $f$ with the same image in $R^2$.

Let $\{x_1, y_1\}, \ldots, \{x_n, y_n\}$ be the $n$ pairs of points of $[0, 1)$ corresponding to the double points

---

F. J. Craveiro de Carvalho is a graduate from Coimbra University, Portugal, where he currently teaches Geometry and Topology. He is a former research student of Stewart A. Robertson, Southampton University, U.K., under whose supervision he wrote a Ph.D. Thesis.

Outside of mathematics his chief interest is literature. He has published some poetry. Although not a good one, he is also a very keen tennis player.

*Supported by Calouste Gulbenkian Foundation-Lisbon, Portugal.

of $f$. We assume that $x_1 < x_2 < \cdots < x_n$ and $x_i < y_i$, for every $i$. It is clear that $f$ has at most $2n$ loops. If they exist, they are given by $f|[x_i, y_i]$, $f|[y_i, x_i + 1]$, for $i = 1, \dots, n$. In this case we say $f$ is *maximally looped*. For example, the curves of Fig. 1 are maximally looped.

The word, $w(f)$, associated with $f$ is written as follows:

Arrange the points $x_i, y_i$, $i = 1, \dots, n$, in their natural order in $R$ to obtain a sequence $z_1 z_2 \dots z_{2n}$. Then

$$w(f) = z_1 z_2 \dots z_{2n}.$$

The idea of associating the word $w(f)$ with $f$ goes back to Gauss [**4**]. He conjectured that if $f$ has $n$ double points, then, for each $i$, the number of symbols between $x_i$ and $y_i$ is even. This conjecture was proved in [**10**]. See also [**8**], [**11**], [**13**].

### 3. A characterization of maximally looped curves.

PROPOSITION 3.1. *Let $f: S^1 \to R^2$ have $n$ double points. Then $f$ has $2n$ loops if and only if $w(f) = x_1 x_2 \dots x_n y_1 y_2 \dots y_n$.*

*Proof.* If $w(f) = x_1 x_2 \dots x_n y_1 y_2 \dots y_n$, then $f$ has $2n$ loops.

On the other hand, suppose that $w(f)$ is not of this form. Then at least one of the $y$'s must be out of order. This can occur in two ways:

(1) there is a pair $\{x_i, y_i\}$ such that $x_1 < y_i < x_n$, or

(2) there are 2 pairs $\{x_i, y_i\}$, $\{x_j, y_j\}$, with $i < j$, such that $y_j$ appears before $y_i$ in $w(f)$, i.e., $x_i < x_j < y_j < y_i$.

In case (2) we do not have $2n$ loops since $f|[x_i, y_i]$ is not a loop ($x_j, y_j \in [x_i, y_i]$ and yet $f(x_j) = f(y_j)$). As far as case (1) is concerned, $i < n$ (since $x_n < y_n$) and $y_i < x_n < y_n < x_i + 1$, thus $f|[y_i, x_i + 1]$ is not a loop.

THEOREM 3.1. *If $f: S^1 \to R^2$ has $n$ double points and more than $n$ loops, then $n$ is odd.*

*Proof.* Since the number of loops is greater than the number of double points, at least one double point gives rise to two loops. We may assume without loss of generality that $\{x_1, y_1\}$ gives rise to the two loops $f|[x_1, y_1]$ and $f|[y_1, x_1 + 1]$. Then

(1) $y_1 \not< x_n$ (otherwise $y_1 < x_n < y_n < x_1 + 1$ and $f|[y_1, x_1 + 1]$ is not a loop) and

(2) $y_i \not< y_1$ (otherwise $x_1 < x_i < y_i < y_1$ and $f|[x_1, y_1]$ is not a loop).

Thus $w(f) = x_1 \dots x_n y_1 z_{n+2} \dots z_{2n}$, that is $z_{n+2}, z_{n+3}, \dots, z_{2n}$ are $y_2, y_3, \dots, y_n$ in some order.

Successive arcs in the sequence

$$(*) \qquad f|(y_1, z_{n+2}), f|(z_{n+2}, z_{n+3}), \dots, f|(z_{2n-1}, z_{2n}), f|(z_{2n}, x_1 + 1)$$

must lie in different components of $R^2 - f([x_1, y_1]$ since $f|(z_i, z_{i+2})$ crosses $f([x_1, y_1])$ transversally at $f(z_{i+1})$, for $i = n + 1, \dots, 2n - 1$, and $z_{2n+1} = x_1 + 1$ (see Fig. 2).

Also by transversality $f|(y_1, z_{n+2})$ and $f|(z_{2n}, x_1 + 1)$ must lie in the same component of $R^2 - f([x_1, y_1])$ (up to orientation, the two possibilities are shown in Fig. 3).

Thus the arcs

$$(**) \qquad f|(y_1, z_{n+2}), f|(z_{n+3}, z_{n+4}), \dots, f|(z_{2n-2}, z_{2n-1}), f|(z_{2n}, x_1 + 1)$$

must lie in one component of $R^2 - f([x_1, y_1])$ and the rest of the arcs

$$(***) \qquad f|(z_{n+2}, z_{n+3}), f|(z_{n+4}, z_{n+5}), \dots, f|(z_{2n-1}, z_{2n})$$

must lie in the other component.

Since there is one more arc in $(**)$ than in $(***)$ the total number, $n$, of arcs in $(*)$ must be odd.

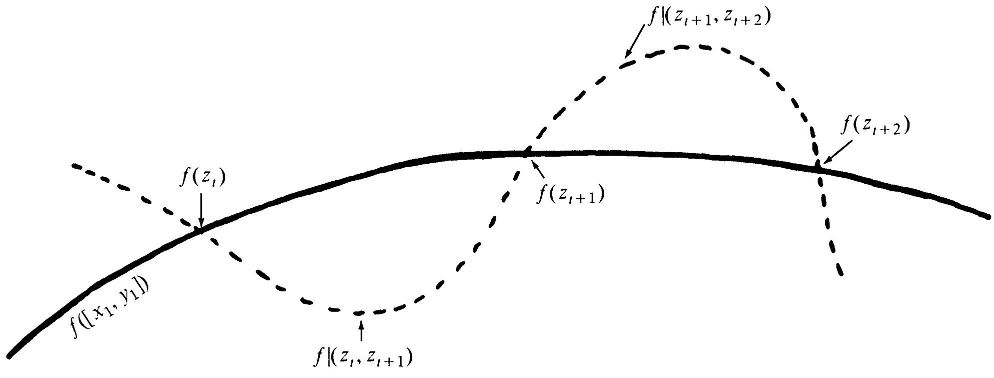



FIG. 2

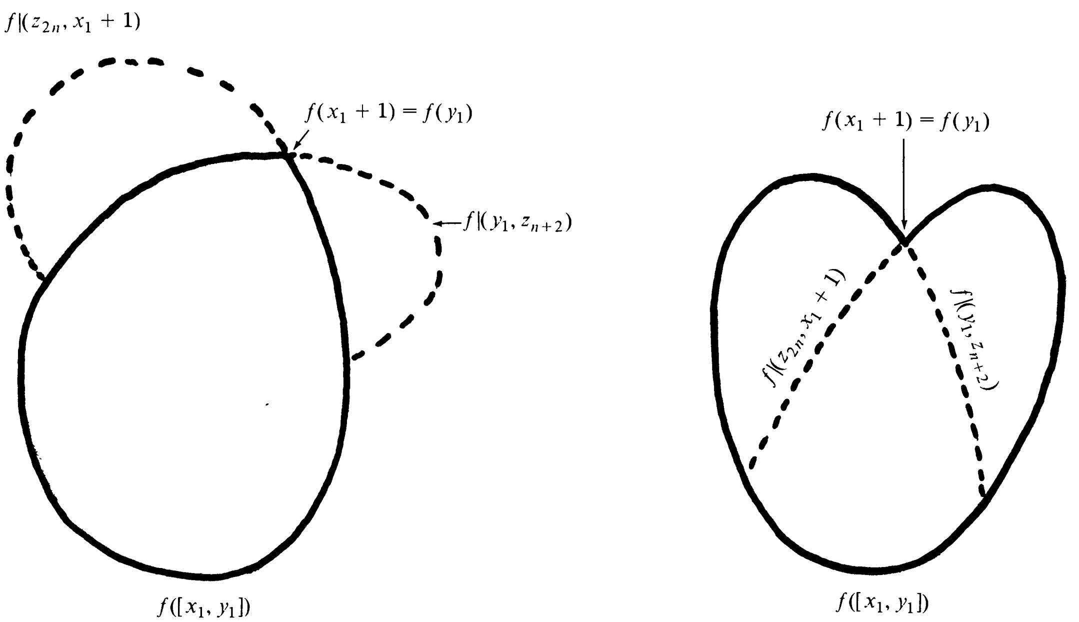



FIG. 3

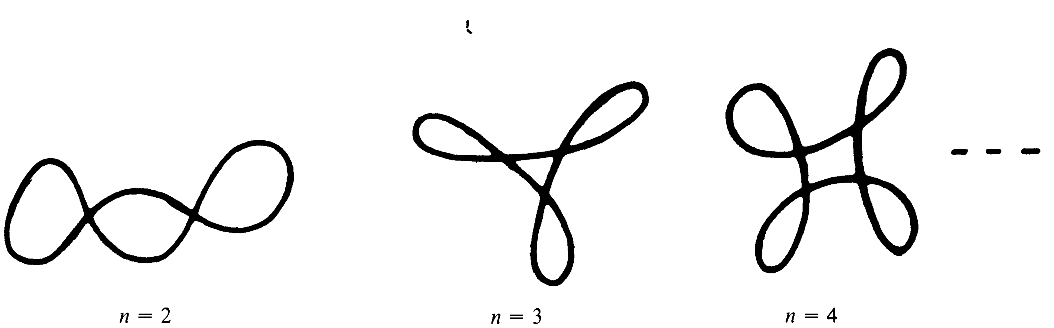



FIG. 4

REMARK. Theorem 3.1 is, in fact, a corollary to Nagy's result. It follows immediately after having established $w(f) = x_1x_2 \dots x_n y_1 z_{n+2} \dots z_{2n}$. However we have preferred to give an independent proof.

An immediate consequence of Theorem 3.1 is

COROLLARY 3.1. *A maximally looped curve must have an odd number of double points.*

It is natural to ask the following question. Suppose $k$ is an integer such that $2 \leqslant k \leqslant n$, if $n$ is even, and $2 \leqslant k \leqslant 2n$, if $n$ is odd. Is there a curve with $k$ loops?

We do not know the complete answer to this question but something can be said. If $k = n \geqslant 2$, then Fig. 4 shows that there are curves with $n$ double points and $n$ loops.

From experimental sketching evidence it seems reasonable to conjecture that if $2 \leqslant k \leqslant n$, then there are closed curves with $k$ loops. However, if $k > n$ that is not the case. For instance, for $n = 3$, if $k > n$ and $f$ has $k$ loops, then $k = 6$. In fact, if $k > n$ there must be a double point giving rise to 2 loops. We shall assume that it is $\{x_1, y_1\}$. Therefore $w(f) = x_1x_2x_3y_1z_5z_6$. Nagy's result implies $z_5 = y_2$ and $z_6 = y_3$. Consequently $f$ has 6 loops. Similarly, if $n = 5$, then the only integers $k > 5$ for which there are closed curves with $k$ loops are $k = 6$ and $k = 10$. The words are

$$x_1x_2x_3x_4x_5y_1y_2y_5y_4y_3 \quad \text{or} \quad x_1x_2x_3x_4x_5y_1y_4y_3y_2y_5, \quad \text{and} \quad x_1x_2x_3x_4x_5y_1y_2y_3y_4y_5.$$

Notice that there is no $f$ with $w(f) = x_1x_2x_3x_4x_5y_1y_4y_5y_2y_3$.

**4. Rotation numbers and double tangents of maximally looped curves.** For some of the ideas we deal with in this section references [3] and [9] are suitable. We start by recalling the notion of *rotation number* of a closed curve. Roughly speaking, it is the total net angle through which the tangent turns while traversing the curve. If $g: R \to R^2$ with period $L$ is an orientation-preserving reparametrization by arc-length of $f$, then the rotation number of $f$ is given by $\int_0^L k(s)\,ds$, where $k(s)$ is the curvature of $g$ at $s$.

If we look again at Fig. 1, we see that for the curves in the top row the rotation number is $4\pi$ or $-4\pi$, depending on the orientation, and is zero for the curves in the bottom row. The next theorem shows that for maximally looped curves these are the only possibilities.

THEOREM 4.1. *If $f: S^1 \to R^2$ is such that $f$ has $2n$ loops, then*

$$|\operatorname{rot} f| = 0 \text{ or } 4\pi,$$

*where* rot $f$ *stands for the rotation number of* $f$.

*Proof.* Without loss of generality we may assume that $f(R)$ lies entirely in one of the closed half-spaces determined by the tangent to $f$ at 0. We shall use a result due to Whitney [14]. He has shown that

$$\operatorname{rot} f = 2\pi[\pm 1 + (n^+ - n^-)],$$

where $n^+$ (resp., $n^-$) is the number of positive (resp., negative) double points. We recall that a double point $\{x, y\}, x < y$, is negative (resp., positive) if the vectors $f'(x), f'(y)$ determine the same orientation as $(1,0), (0,1)$ in $R^2$ (resp., the opposite orientation).

We modify slightly the loop $f: [x_1, y_1] \to R^2$ near $x_1$ and $y_1$ so that we obtain a $C^1$ embedding $g: S^1 \to R^2$ (see Fig. 5).

Next we consider the two possible orientations of $g(S^1)$.

(1) $S^1$ is given the usual counterclockwise orientation and $g(S^1)$ inherits the orientation induced by $g$.

(2) $g(S^1)$ is oriented as the boundary of a closed disc with the standard orientation in $R^2$ [7].

If the two orientations of $g(S^1)$ agree, then each time $f: [y_1, x_1 + 1] \to R^2$ crosses $g(S^1)$ from the outside (resp., inside) to the inside (resp., outside) of $g(S^1)$ we have a negative (resp., positive)

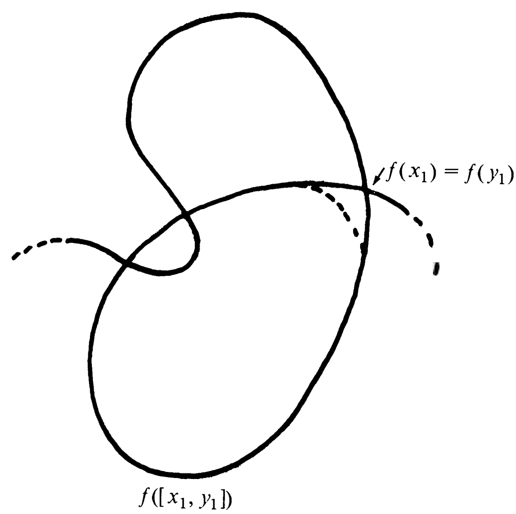


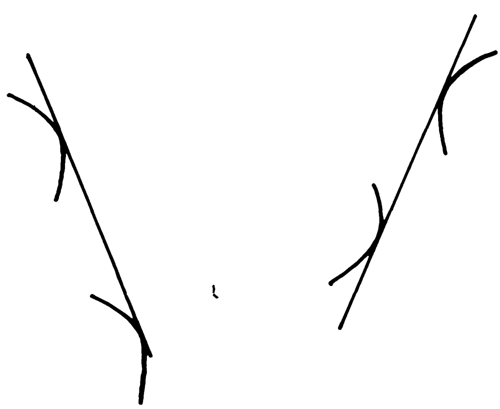

FIG. 5

double point. If the two orientations disagree, then we have positive and negative double points, respectively. Using Whitney's formula above we conclude that $|\mathrm{rot} f| = 0$ or $4\pi$.

*Note*. The reader familiar with the notion of regular homotopy and the Whitney-Graustein theorem will see that up to orientation and regular homotopy the only maximally looped curves are those we show in Fig. 1 [**14**].

The curves in the top row of Fig. 1 have exactly $n$ double tangents. Theorem 4.2 states that this is always true for maximally looped $C^4$ curves without inflection points. Such curves are examples of *double ovaloids* [**12**].

Let us suppose that $f: R \to R^2$ is $C^4$. Denoting the tangent to $f$ at $x \in R$ by $T_x$, we say that a pair $\{x, y\} \subset [0,1)$ of distinct points is a double tangent if $T_x = T_y$. It is *regular* if the curvature of $f$ at $x$ and $y$ is nonzero. If a double tangent is regular, then it is of type (a) or (b) (see Fig. 6).



(a) (b)

FIG. 6

We say the double tangent is *external* in case (a) and *internal* in case (b).

An inflection point $x$ of $f$ is *regular* if the curvature has nonzero derivative at $x$.

Suppose that $f: R \to R^2$ is self-transverse, and that the double tangents and inflection points are regular. Denote by $E$, $I$, $n$ and $F$ the numbers of external double tangents, internal double

tangents, double points, and inflection points, respectively. It is known ([**2**], [**5**]) that

$$E = I + n + \frac{1}{2}F.$$

THEOREM 4.2. *If $f\colon S^1 \to R^2$ has $2n$ loops and no inflection points, then $f$ has exactly $n$ double tangents.*

*Proof.* If $f$ has no inflection points, then its curvature is always positive or always negative. Since the curvature of $f$ is the same as the curvature of any orientation-preserving reparametrization of $f$ by arc-length, it follows that $\operatorname{rot} f$ is either $> 0$ or $< 0$. Therefore $|\operatorname{rot} f| = 4\pi$. Halpern has shown in [**6**] that

$$4\pi^2 I \leqslant (|\operatorname{rot} f| - 4\pi)(2|\operatorname{rot} f| - 2\pi).$$

Consequently $I = 0$. The result now follows from the equality $E = I + n + \frac{1}{2}F$.

**5. Final comments.** There are a few directions in which we could try to pursue the ideas we dealt with in this note. We mention two.

(1) What results hold if triple points, quadruple points, etc., are allowed?

(2) Every closed curve without double points separates $R^2$ in 2 regions. This result was used in our work but it is not true for an arbitrary surface. What happens if $R^2$ is replaced by a torus, for instance? The situation is then quite different. While, for example, there is no $f\colon S^1 \to R^2$ such that $w(f) = x_1x_2y_1y_2$, it is possible to obtain $f\colon S^1 \to$ Torus with word $x_1x_2y_1y_2$. Fig. 7 shows how.

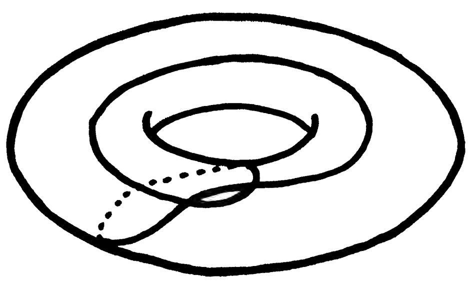


FIG. 7

We end up this note by calling the interested reader's attention to the article by J. W. Bruce and P. J. Giblin, "Generic Geometry", which appeared in this MONTHLY, vol. 90, 529–545. Their techniques and results are a different approach to global problems in the theory of smooth plane curves.

## References


**1.** J. Eells, Singularities of Smooth Maps, Thomas Nelson and Sons, 1968.

**2.** Fr. Fabricius-Bjerre, On the double tangents of plane closed curves, Math. Scand., 11 (1962) 113–116.

**3.** M. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, 1976.

**4.** C. F. Gauss, Werke, Teubner, Leipzig, 1900, 282–286.

**5.** B. Halpern, Global theorems for closed plane curves, Bull. Amer. Math. Soc., 76 (1970) 96–100.

**6.** ______, An inequality for double tangents, Proc. Amer. Math. Soc., 76 (1979) 133–139.

**7.** M. W. Hirsch, Differential Topology, Springer-Verlag, 1976.

**8.** M. L. Marx, The Gauss realizability problem, Proc. Amer. Math. Soc., 22 (1969) 610–613.

**9.** R. S. Millman and G. Parker, Elements of Differential Geometry, Prentice-Hall, 1977.

**10.** J. V. Sz. Nagy, Uber ein topologisches Problem von Gauss, Math. Z., 26 (1927) 579–592.

**11.** C. J. Titus, A theory of normal curves and some applications, Pacific J. Math., 10 (1960) 1083–1096.

**12.** L. F. Tóth, Regular Figures, Pergamon Press, 1964.

**13.** L. B. Treybig, A characterization of the double point structure of the projection of a poligonal knot in regular position, Trans. Amer. Math. Soc., 130 (1968) 223–247.

**14.** H. Whitney, On regular closed curves in the plane, Compositio Math., 4 (1937) 276–284.

C E N T E R   S E C T I O N
(Vol. 92, No. 3, March 1985)

**Telegraphic Reviews**

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S, P, L.** Mathematics Today: Twelve Informal Essays. Ed: Lynn Arthur Steen. Springer-Verlag, 1984, viii + 367 pp, $28. [ISBN: 0-387-90305-4] Third printing of volume of essays first published in 1978 (TR, March 1979; Extended Review, December 1979).

**General, P, L.** Research Briefings 1984. National Academy Press, 1984, vii + 116 pp, (P). [ISBN: 0-309-03495-7] Nine research reports prepared in 1984 for the Office of Science and Technology Policy (OSTP). Includes papers on computer architecture, and on information technology in precollege education, plus others in biology, chemistry, and physics. LAS

**General, S*, L.** Mathematical Applications of Electronic Spreadsheets. Deane Arganbright. McGraw-Hill, 1985, x + 165 pp, $16.95 (P). [ISBN: 0-07-002429-4] Examples of how to set up electronic spreadsheets such as Lotus, VisiCalc, or Multiplan to execute a variety of mathematics techniques: Newton's method, Taylor polynomials, eigenvalues, continued fractions, Bayes' rule, simplex algorithm, and much more. A convincing demonstration of the programming capabilities of spreadsheets, as well as an innovative and interesting means of introducing algorithms into mathematics instruction. LAS

**General, T(15-17: 2, 3).** Corso di matematica superiore, Volume Terzo. Vladimir Ivanovič Smirnov. Editori Riuniti, 1982. Parte prima, 316 pp; Parte seconda, 684 pp. Italian translation of third volume of Russian text for course in advanced mathematics. Covers linear algebra, groups and group representations, elementary complex analysis, applications to physics, linear differential equations, special functions. No exercises. KS

**General, S(13).** Beyond Problem Solving and Comprehension: An Exploration of Quantitative Reasoning. Arthur Whimbey, Jack Lochhead. Franklin Inst Pr, 1984, v + 392 pp, $14.95 (P). [ISBN: 0-89168-051-9] A follow-up to the authors' earlier work Problem Solving and Comprehension (TR, June-July 1984). Intended to help students develop analytical reasoning skills by working on problems in problem solver and listener pairs. Solutions or answers provided for all problems. JRG

**General, S*(13-18), L**.** Puzzles and Paradoxes: Fascinating Excursions in Recreational Mathematics. T.H. O'Beirne. Dover, 1984, xiv + 238 pp, $4.95 (P). [ISBN: 0-486-24613-2] Dover republication of a work published by Oxford University Press in 1965. Material originally appeared in the periodical New Scientist in 1961 and 1962. The best of its kind. Clear, detailed, masterly presentations of some familiar and some not-so-familiar puzzles and paradoxes. Superb postscript with traces of history and suggestions for further reading. Excellent references, many to first-hand accounts. This republication is a real service to devotees of recreational mathematics and a steal at the price. JK

**General, L.** Lecture Notes in Mathematics-501-1000: An Index and Other Useful Information. Ed: A. Dold, B. Eckmann. Springer-Verlag, 1984, 108 pp, (P). Title, author, and subject indices, plus advice to prospective LNM authors. This well-known and much copied series has produced 1000 volumes in 20 years (1964-1984). LAS

**Elementary, S.** Short-cut Math. Gerard W. Kelly. Dover, 1984, 112 pp, $2.95 (P). [ISBN: 0-486-24611-6] Unabridged and slightly revised republication of a Sterling edition published in 1969; contains about 150 shortcuts for arithmetic operations. JNC

**Elementary, T(13: 1).** Beginning Algebra for College Students, Third Edition. Karl J. Smith, Patrick J. Boyle. Brooks/Cole, 1984, xi + 313 pp, $25.95. [ISBN: 0-534-02779-2] Designed "to appeal to a variety of individuals and to allay (students') fear of math." Emphasis on applications and word problems as well as drill. Wonderful quotes from literature. New in third edition: polynomials, factoring and rational expressions before graphs and systems of equations. (First Edition, TR, January 1977.) MA

**Precalculus, T(13: 1).** Fundamentals of College Algebra and Trigonometry. Robert Ellis, Denny Gulick. HarBrace J, 1984, xi + 448 pp, $22.95. [ISBN: 0-15-529350-8] Identical to first eight chapters of College Algebra and Trigonometry, Second Edition (TR, December 1984). Interesting applied problem with photograph motivates each chapter. Ample selection of exercises--many word problems. New in Second Edition: trigonometry begins with right triangles, includes polar coordinates. MA

**Education, P, L*.** American Perspectives on the Fifth International Congress on Mathematical Education (ICME 5). Ed: Warren Page. MAA, 1985, viii + 134 pp, (P). [ISBN: 0-88385-055-9] Brief communications from each of the American delegates to the August 1984 Fifth International Congress on Mathematics Education in Adelaide, Australia provide diverse and often contrasting perspectives on the presentations and experiences of the Congress. Technology, problem solving, contests and societal issues supplement general mathematics education themes from around the world. LAS

**Education, L.** Cumulative Index, The Arithmetic Teacher, 1974-1983: Volumes 21-30. NCTM, 1984, 34 pp, (P). Author and subject index to the third decade of The Arithmetic Teacher. LAS

**History, L*.** Babbage's Calculating Engines: A Collection of Papers by Henry Prevost Babbage. Henry Prevost Babbage. Reprint Ser. for the History of Computing, V. II. Tomash, 1982, 30 + 359 pp, $45. [ISBN: 0-938228-04-8] A collection of writings of Charles Babbage and contemporary commentary on his work. A substantial portion of this volume is devoted to excerpts from the autobiographical 1889 edition. JRG

**History, P*, L.** Mathematical Astronomy in Copernicus's De Revolutionibus. N.M. Swerdlow, O. Neugebauer. Stud. in History of Math. & Physical Sci., No. 10. Springer-Verlag, 1984, $78 set [ISBN: 0-387-90939-7]. Part 1, xvi + 537 pp; Part 2, v + 172 pp. Analysis and exposition of De revolutionibus in relation to the methodology of Plotemy's Almagest, completing intellectually the history of ancient astronomy begun by Neugebauer in Volume 1 of this series. LAS

**History, P, L.** Families of Curves and the Origins of Partial Differentiation. Steven B. Engelsman. Math. Stud., V. 93. Elsevier Sci, 1984, ix + 238 pp, $29 (P). [ISBN: 0-444-86897-6] Historical study of the evolution of the concept of partial differentiation in the early 18th century. This "unknown world," an "entirely new way of calculation," emerged from studies of families of curves by the Bernoullis, Leibniz, and Euler. Contains translations of two previously unpublished manuscripts by Euler and Nicolaus I. Bernoulli. LAS

**Foundations, T(15-16: 1), L.** Logic and Structure, Second Edition. Dirk van Dalen. Universitext. Springer-Verlag, 1983, x + 207 pp, $14 (P). [ISBN: 0-387-12831-X] Undergraduate course in mathematical logic. Uses natural deduction and covers main results of model theory. No recursion theory. Only change from First Edition (TR, March 1981) is addition of chapter on intuitionistic logic. KS

**Foundations, T(15-17: 1), S, L.** Introduction to Set Theory, Second Edition, Revised and Expanded. Karel Hrbacek, Thomas Jech. Pure & Appl. Math., V. 85. Dekker, 1984, ix + 247 pp, $39.75. [ISBN: 0-8247-7074-9] A nice expansion of a fine text. One new section deals with sets of real numbers with applications to real analysis, and another includes substantial results about large cardinals, e.g., "Every measurable cardinal is strongly inaccessible." The proofs are both elegant and readable. (First Edition, TR, November 1978.) MU

**Foundations, T(15-16: 1), L.** Mathematical Logic. H.-D. Ebbinghaus, J. Flum, W. Thomas. Undergrad. Texts in Math. Springer-Verlag, 1984, ix + 216 pp, $28. [ISBN: 0-387-90895-1] Translation and revision of 1978 German text for undergraduate logic course. Formal presentation. Covers predicate logic using sequent calculus, beginning model theory, incompleteness using register machines. Introduces second-order and infinitary logics. Short exercise sets at ends of sections. No mathematical prerequisites other than maturity. KS

**Foundations, P.** Surveys in Set Theory. Ed: A.R.D. Mathias. London Math. Soc. Lect. Note Ser., No. 87. Cambridge U Pr, 1983, vii + 247 pp, $27.50 (P). [ISBN: 0-521-27733-7] Expository papers on forcing, morasses and descriptive set theory plus two research articles. Most papers based on presentations at 1978 Cambridge summer school in set theory. KS

**Foundations, S(16-18), P.** Axiomatic Set Theory. Ed: James E. Baumgartner, Donald A. Martin, Saharon Shelah. Contemp. Math., V. 31. AMS, 1984, x + 259 pp, $25 (P). [ISBN: 0-8218-5026-1] This book contains the proceedings of the June 1983 AMS Research Conference covering a vast range of topics, from the history of set theory (Russell on Cantor) through the spectrum of current research. Anyone interested in set theory will be fascinated by several of these papers. MU

**Foundations, T(18: 1), P.** Subrecursion: Functions and Hierarchies. H.E. Rose. Logic Guides, No. 9. Clarendon Pr, 1984, xiii + 191 pp, $35. [ISBN: 0-19-853189-3] Subrecursive functions are relatively simple computable functions. Hierarchies, sequences of classes of functions of increasing complexity, are used to classify and study these functions. This monograph focuses on relationships between various hierarchies and metamathematical properties of formal systems based on classes of subrecursive functions. Includes open problems, exercises. KS

**Graph Theory, S(16-18), P.** Selected Topics in Graph Theory 2. Ed: Lowell W. Beineke, Robin J. Wilson. Academic Pr, 1983, x + 301 pp, $60. [ISBN: 0-12-086202-6] Eight expository surveys on aspects of pure graph theory. Some comprehensible by undergraduates. Includes Eulerian, perfect, infinite and random graphs, automorphism groups, extremal graph theory, graphs and partially ordered sets,

graphs and games. Extensive bibliographies. KS

**Combinatorics, S(15-16), L.** Notes on Introductory Combinatorics. George Pólya, Robert E. Tarjan, Donald R. Woods. Prog. in Comp. Sci., No. 4. Birkhauser Boston, 1983, 192 pp, $9.95 (P). [ISBN: 0-8176-3123-2] Notes prepared by Woods based on course taught by Pólya and Tarjan at Stanford in 1978. Good reference for instructors of discrete mathematics or combinatorics. Covers generating functions, inclusion-exclusion, Pólya's theory, recurrence relations, Ramsey theory, matchings, graphs. Includes homework problems and examinations with solutions, references. KS

**Combinatorics, T**(15: 1), S*, L*.** Introduction to Combinatorial Theory. R.C. Bose, B. Manvel. Wiley, 1984, ix + 237 pp, $29.95. [ISBN: 0-471-89614-4] An excellent introductory text. Includes introductory material on counting, graphs, finite fields and geometries, Latin squares, error-correcting codes, block designs, problems of choice and optimization. Readable. Good collection of exercises. CEC

**Combinatorics, P.** Lecture Notes in Computer Science-171: Logic and Machines: Decision Problems and Complexity. Ed: E. Börger, G. Hasenjaeger, D. Rödding. Springer-Verlag, 1984, vi + 456 pp, $20 (P). [ISBN: 0-387-13331-3] Proceedings from the symposium "Rekursive Kombinatorik" held at the Institut für Mathematische Logik und Grundlagenforschung der Universität Münster, May 23-28, 1983. JAS

**Number Theory, S(18), P.** Classical and Quantum Models and Arithmetic Problems. Ed: David V. and Gregory V. Chudnovsky. Lect. Notes in Pure & Appl. Math., V. 92. Dekker, 1984, xi + 461 pp, $79.50 (P). [ISBN: 0-8247-1825-9] A series of papers devoted to various mathematical aspects of dynamic systems connected with or arising from number theory. The result of a seminar entitled "Exact Solutions of Classical and Quantum Models and Parallel Arithmetic Problems," held at Columbia University in 1980-81. CEC

**Number Theory, P.** Lecture Notes in Mathematics-1068: Number Theory, Noordwijkerhout 1983. Ed: H. Jager. Springer-Verlag, 1984, v + 296 pp, $16 (P). [ISBN: 0-387-13356-9] A collection of 24 papers from Journées Arithmétiques which took place July 11-15, 1983 at Noordwijkerhout in the Netherlands. CEC

**Number Theory, P.** Lecture Notes in Mathematics-1046: Algebraic K-Theory, Number Theory, Geometry and Analysis. Ed: A. Bak. Springer-Verlag, 1984, ix + 464 pp, $21.50 (P). [ISBN: 0-387-12891-3] Proceedings of an international conference held at Bielefeld, West Germany, July 26-30, 1982. JAS

**Linear Algebra, S(15-16).** Mathematics for Econometrics, Second Edition. Phoebus J. Dhrymes. Springer-Verlag, 1984, x + 150 pp, $19.80 (P). [ISBN: 3-087-90988-5] A linear algebra handbook, rather than a text. New in this edition are discussions of singular value decomposition and issues of stability of the general linear structural econometric model. (First Edition, TR, April 1979.) JRG

**Linear Algebra, T(14: 1).** Introduction to Linear Algebra. Roger C. McCann. HarBrace J, 1984, ix + 419 pp, $26.95. [ISBN: 0-15-543001-7] Standard sophomore-level linear algebra text; rather short on exercises and illustrations. Applications to physics, computer graphics, differential equations, are included. Some numerical methods. JRG

**Algebra, S(18), P.** Inverse Semigroups. Mario Petrich. Wiley, 1984, x + 674 pp, $62.50. [ISBN: 0-471-87545-7] A self-contained and comprehensive treatment of inverse semigroups (regular with commuting idempotents) from the beginning to "the boundaries of our current knowledge." Includes such topics as special classes, representations, hulls, E-unitary, monogenic, bisimple, $\omega$-regular semigroups and discussion of varieties, amalgamation, and trace. Exercises, open questions, bibliography, index. JS

**Algebra, P.** Lecture Notes in Mathematics-1041: Lie Group Representations II. Ed: R. Herb, et al. Springer-Verlag, 1984, ix + 340 pp, $15.50 (P). [ISBN: 0-387-12715-1] Papers from November during the 1982-1983 special year at the University of Maryland. JAS

**Algebra, P.** Relative Invariants of Rings: The Noncommutative Theory. F. Van Oystaeyen, A. Verschoren. Pure & Appl. Math., V. 86. Dekker, 1984, xii + 286 pp, $59.75. [ISBN: 0-8247-7281-4] The non-commutative counterpart of the authors' earlier book for commutative rings, this is an advanced treatment beginning with Krull domains and maximal orders and continuing with chapters on relative Picard groups, Azumaya algebras, and construction of orders. Extensive references, index. JS

**Algebra, P*.** Algebra and Its Applications. Ed: H.L. Manocha, J.B. Srivastava. Lect. Notes in Pure & Appl. Math., V. 91. Dekker, 1984, xii + 395 pp, $59.50 (P). [ISBN: 0-8247-7165-6] Collection of 48 papers on topics of current research, sponsored by or presented at the International Symposium on Algebra and its Applications, New Delhi, 1981. Few papers discuss actual applications; nevertheless a valuable summary of current research in ring theory, algebraic geometry, coding theory and graph theory. JRG

**Algebra, T*(17: 1), S*, P*, L*.** Galois Theory. Harold M. Edwards. Grad. Texts in Math., No. 101. Springer-Verlag, 1984, xiii + 152 pp, $22. [ISBN: 0-387-90980-X] An historical approach to the work of Galois on the solvability of equations by radicals. Includes some of the work of Galois' antecedents, a translation of Galois' memoir and an exceptional collection of exercises. Exciting reading! CEC

**Algebra, S(18), P.** Wittrings. Manfred Knebusch, Manfred Kolster. Aspects of Math. Heyden & Son, 1982, xi + 96 pp, $14 (P). [ISBN: 3-528-08512-6] An introduction to some parts of the algebraic theory of quadratic forms which have developed recently. Presents the structure theory of Wittrings and reduced Wittrings over fields. No exercises, bibliography. CEC

**Algebra, T*(15-17: 2), S, L*.** Abstract Algebra: A Computational Approach. Charles C. Sims. Wiley, 1984, xv + 491 pp, $32.95. [ISBN: 0-471-09846-9] Material for a year's course--or more. Uses APL and is rendered self-contained via an appendix and exercises (of the form "express this in APL") throughout. The algebra is substantial and pretty standard but algorithms for doing it are added. The choice of APL would appear to make this addition not only feasible but worthwhile. An instructor's manual and a manual with APL library routines are also available. JAS

**Calculus, S(13).** O Conceito de Derivacão. V. Boltiansky. MIR, 1983, 56 pp, (P). Portuguese translation of Russian booklet on differentiation. Uses object falling with air resistance to motivate derivative. Discusses maxima and minima and differential equations associated with circuits, radioactive decay, harmonic oscillations. KS

**Calculus, S(13).** Differential and Integral Calculus. R.L. Wallis. New Math. Lib., V. 11. Van Nostrand Reinhold, 1984, viii + 153 pp, $13.95 (P). [ISBN: 0-442-30579-6] Part of a British series designed to cover mainstream topics in mathematics. A concise, rigorous and tightly organized presentation of the major topics of elementary calculus including partial differentiation, maxima and minima, error analysis, and Taylor series. Exercises. An excellent way to review calculus. JNC

**Real Analysis, T(15), S, L.** Convex Analysis: An Introductory Text. Jan van Tiel. Wiley, 1984, viii + 125 pp, $21.95. [ISBN: 0-471-90263-2] An introduction to convex functions set first in the context of the real line, then in normed linear spaces. AWR

**Real Analysis, S(17-18), P.** Toeplitz Forms and Their Applications. Ulf Grenander, Gabor Szegö. Chelsea, 1984, ix + 245 pp, $16.95. [ISBN: 0-8284-0321-X] Reprint (on long-life paper) of original 1955 edition. An exposition of the theory of Toeplitz forms followed by applications to analytic functions, probability theory, and mathematical statistics. JS

**Complex Analysis, S(18), P, L.** Uniform Algebras. Theodore W. Gamelin. Chelsea, 1984, xiii + 269 pp, $17.95. [ISBN: 0-8284-0311-2] Corrected reprint of 1969 original edition (TR, January 1970), supplemented by a short annotated list of new references. "The central problem in the field of uniform algebras is to decide whether a given complex-valued function can be uniformly approximated by members of a prescribed algebra of functions." LAS

**Differential Equations, P.** Lecture Notes in Mathematics-1047: Fluid Dynamics. Ed: H. Beirã da Veiga. Springer-Verlag, 1984, vii + 187 pp, $9.50 (P). [ISBN: 0-387-12893-X] Lectures from the Centro Internazionale Matematico Estivo held in Varenna, Italy, August 22-September 1, 1982. JAS

**Differential Equations, P.** Numerical Solution of Partial Differential Equations: Theory, Tools and Case Studies. Ed: D.P. Laurie. ISNM 66. Birkhauser Boston, 1983, 341 pp, $29.95. [ISBN: 3-7643-1561-X] Proceedings of the summer seminar series held at CSIR, Pretoria, South Africa, February 8-10, 1982. JAS

**Differential Equations, T(16: 2).** Métodos de solución de las ecuaciones reticulares. A.A. Samarski, E.S. Nikolaev. MIR, 1982. Tomo I, 334 pp; Tomo II, 448 pp. Spanish translation of Russian monographs on methods for solving difference equations. Covers direct and iterative methods. KS

**Differential Equations, P.** Lecture Notes in Mathematics-1057: Bifurcation Theory and Applications. Ed: L. Salvadori. Springer-Verlag, 1984, vii + 233 pp, $10 (P). [ISBN: 0-387-12931-6] Four sets of lectures (by J.K. Hale, J.J. Dvistermaat, G. Iooss, S. Busenberg, respectively) from the CIME international summer course at Montecatini, Italy, 1983. Includes additional lectures given there by W.S. Loud and A. Vanderbauwhede. DA

**Differential Equations, T(18), P.** An Introduction to Infinite Dimensional Dynamical Systems--Geometric Theory. Jack K. Hale, Luis T. Magalhães, Waldyr M. Oliva. Appl. Math. Sci., No. 47. Springer-Verlag, 1984, 195 pp, $18 (P). [ISBN: 0-387-90931-1] An approach to the theory of infinite dimensional dynamical systems by analogy with the theory of finite dimensions, emphasizing retarded functional differential equations. An extensive appendix provides an introduction to the homotopy index theorem in noncompact spaces. LAS

**Differential Equations, T(17-18: 1, 2).** Mathematical Methods for Wave Phenomena. Norman Bleistein. Computer Sci. & Appl. Math. Academic Pr, 1984, xv + 341 pp, $55. [ISBN: 0-12-105650-3] Designed for use as a textbook in a methods of applied mathematics course. Presumes a basic background in real and complex analysis, differential equations, and linear algebra. The final two chapters discuss asymptotic techniques and inverse methods. AO

**Partial Differential Equations, P.** Nonlinear Partial Differential Equations in Applied Science; Proceedings of the U.S.-Japan Seminar, Tokyo, 1982. Ed: Hiroshi Fujita, Peter D. Lax, Gilbert Strang. Math. Stud., V. 81. Elsevier Sci, 1983, xviii + 457 pp, $60 (P). [ISBN: 0-444-86681-7] Proceedings of the U.S.-Japan Seminar held in Tokyo in July 1982. JAS

**Partial Differential Equations, S(18), P.** Lecture Notes in Mathematics-1058: Continuous and Discrete Dynamics near Manifolds of Equilibria. Bernd Aulbach. Springer-Verlag, 1984, ix + 142 pp, $8.50 (P). [ISBN: 0-387-13329-1] Studies the stationary solutions of nonlinear ordinary differential and difference equations; focuses on the more difficult case of manifolds of constant solutions, rather than isolated stationary points. Describes conditions under which a solution converges to a given point on the manifolds of stationary solutions. Applies these results to the dynamics of the distribution of genotypes within a population with overlapping generations. YN

**Partial Differential Equations, S**(18), P*.** Partial Differential Equations and Dynamical Systems. Ed: W.E. Fitzgibbon III. Research Notes in Math., V. 101. Pitman, 1984, 366 pp, $24.95 (P). [ISBN: 0-273-08644-8] Lively collection of well written research and expository papers. Each essay focuses on modelling a real phenomenon with partial differential equations, and analyzing the properties of the solutions of these equations (smoothness, stability, etc.). Requires a modest background in differential geometry, functional analysis, and phase plane methods. Example: bifurcation problems in the buckling of cylindrical shells. YN

**Partial Differential Equations, S(17), P.** A Layering Method for Viscous, Incompressible $L_p$ Flows Occupying $R^n$. A. Douglis, E.B. Fabes. Research Notes in Math., V. 108. Pitman, 1984, 171 pp, $19.95 (P). [ISBN: 0-273-08650-2] Constructs solutions for the Navier-Stokes equations. The layering method involves two steps. First, smoothing the data by convolution with a smooth kernel (approximate Dirac distribution), which allows for polynomial approximations in the second step. Second, solving a related "layer" equation. Both steps are then performed over consecutive time intervals (the "layers"). Technical and occasionally laconic, but interesting and readable. YN

**Partial Differential Equations, T(17: 1), S*, P.** Lecture Notes in Mathematics-1072: Global Solutions of Reaction-Diffusion Systems. Franz Rothe. Springer-Verlag, 1984, v + 216 pp, $11 (P). [ISBN: 0-387-13365-8] Lucid study of nonlinear parabolic problems. Makes realistic regularity hypotheses, reflecting real phenomena, and then establishes the necessary mathematics to obtain existence, uniqueness, and properties of the solutions--rather than assuming nice arbitrary conditions that just make the math work. Constructs criteria to distinguish uniform a priori bounds from explosion situations. Very readable in spite of its technical nature and sometimes odd terminology. Applications: nerve pulse along an axon, nuclear reactor, etc. YN

**Numerical Analysis, P.** Lecture Notes in Mathematics-1071: Padé Approximation and its Applications, Bad Honnef 1983. Ed: H. Werner, H.J. Bünger. Springer-Verlag, 1984, vi + 264 pp, $13.50 (P). [ISBN: 0-387-13364-X] Proceedings of a March 1983 symposium at the Physik-Zentrum in Bad Honnef. 18 papers dealing with functions approximated by means of rational expressions, including accelerated convergence of continued fractions and algorithms for multivariable rational interpolation. LAS

**Numerical Analysis, T(15: 1, 2), L.** Computational Mathematics: An Introduction to Numerical Approximation. T.R.F. Nonweiler. Ser. in Math. & Its Applic. Halsted Pr, 1984, 431 pp, $59.95. [ISBN: 0-470-27472-7] Discursive in nature, leaving much work to the comprehensive, well-written problem sets. Provides hints and answers. Computational approach. Topics: computational arithmetic, summing series, continued fractions, function approximation by interpolation, finding real zeros of functions of one variable, numerical quadrature. Appendix gives BASIC numerical subroutines. DA

**Numerical Analysis, T(16-18: 1), P, L.** Lecture Notes in Computer Science-165: Large Sparse Numerical Optimization. Thomas F. Coleman. Springer-Verlag, 1984, v + 105 pp, $8.50 (P). [ISBN: 0-387-12914-6] Notes from a graduate course. Begins with a survey of algorithms for large sparse linear problems (square systems of linear equations, overdetermined systems, and linear programming) then focuses on nonlinear problems. Concludes with a chapter on large sparse quadratic programming problems. A good review of the state of the art. AO

**Numerical Analysis, P.** Approximation Theory IV. Ed: C.K. Chui, L.L. Schumaker, J.D. Ward. Academic Pr, 1983, xvii + 785 pp, $50. [ISBN: 0-12-174580-5] Seven survey papers and 74 research notes presented at a January 1983 international symposium at Texas A&M. Concludes with a lengthy bibliography on Bernstein polynomials and applications covering 1955-1982. LAS

**Numerical Analysis, T(16-17: 1, 2), L.** Sparse Matrix Technology. Sergio Pissanetzky. Academic Pr, 1984, xiii + 321 pp, $55. [ISBN: 0-12-557580-7] An introductory survey of the special numerical techniques available for sparse matrix problems. Covers the solution of linear algebraic equations, eigenanalysis, and related topics. FORTRAN implementations of some algorithms are given. AO

**Numerical Analysis, P.** Lecture Notes in Mathematics-1066: Numerical Analysis. Ed: David F. Griffiths. Springer-Verlag, 1984, xi + 275 pp, $14 (P). [ISBN: 0-387-13344-5] Proceedings of the Tenth Dundee Biennial Conference on Numerical Analysis held June 28-July 1, 1983. Contains fifteen papers by the invited speakers that cover a broad range of topics. AO

**Functional Analysis, P.** Lecture Notes in Mathematics-1070: Interpolation Spaces and Allied Topics in Analysis. Ed: M. Cwikel, J. Peetre. Springer-Verlag, 1984, 239 pp, $11 (P). [ISBN: 0-387-13363-1] Proceedings of the conference held in Lund, Sweden, August 21-September 1, 1983. JAS

**Functional Analysis, S(16-17).** Teoremi e problemi dell'analisi funzionale. Aleksandr Kirillov, Aleksej Gvišiani. MIR, 1983, 358 pp. Italian translation of Russian problem book on functional analysis. Summary of major definitions and theorems, 851 problems, hints or solutions for all problems. Covers measure and integration, linear spaces and operators, Fourier transforms, spectral theory. KS

**Functional Analysis, S(16-18).** Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings. Kazimierz Goebel, Simeon Reich. Pure & Appl. Math., V. 83. Dekker, 1984, ix + 170 pp, $35. [ISBN: 0-8247-7223-7] From lectures by Goebel in a nonlinear functional analysis seminar at U.S.C. Relates nonexpansive and holomorphic mappings on complex, infinite-dimensional Banach spaces. Uses only metric methods and assumes only modest background in functional analysis and complex analysis. DA

**Analysis, T*(18: 4), S*, P.** Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions. Sigurdur Helgason. Pure & Appl. Math., V. 113. Academic Pr, 1984, xix + 654 pp, $39.50. [ISBN: 0-12-338301-3] Treats harmonic analysis, integral geometry, and differential operators on homogeneous spaces: inversion theorems for general Fourier transforms (assumes the ordinary one), Radon transforms in any dimension, and spherical transforms. Local and global solvability, and eigenfunctions of invariant differential operators. Representations of Lie groups. Gives a few exercises, with solutions. Mentions applications. Claims to be a reference on integral geometry but ignores Gel´fand´s recent work where, e.g., the difference in Radon inversion formulae for even or odd dimensions is related to the different homotopy types of the unit sphere. Nevertheless, contains a lot of good mathematics. YN

**Analysis, P.** Lecture Notes in Mathematics-1061: Séminaire de Théorie du Potentiel Paris, No. 7. F. Hirsch, G. Mokobodzki. Springer-Verlag, 1984, iv + 281 pp, $14 (P). [ISBN: 0-387-13338-0]

**Analysis, P.** Nondiscrete Induction and Iterative Processes. F-A. Potra, V. Ptak. Research Notes in Math., V. 103. Pitman, 1984, 207 pp, $21.95 (P). [ISBN: 0-273-08627-8] Surveys papers devoted to nondiscrete mathematical induction and sketches the present state of the method. Applications to numerical analysis, factorization theorems, transitivity in operator algebras, stability of openness for linear mappings and exactness of sequences, interpolation spaces and iterative construction of solutions to nonlinear partial differential equations. DA

**Analysis, P.** Lecture Notes in Mathematics-1065: Padé Approximants for Operators: Theory and Applications. Annie Cuyt. Springer-Verlag, 1984, ix + 138 pp, $8 (P). [ISBN: 0-387-13342-9] Generalizes the notion of Padé approximation for real functions to nonlinear operators between Banach spaces. Includes existence theorems, covariance properties, recurrence relations, comparison with other generalizations, and applications to solutions of nonlinear equations. RM

**Analysis, P.** Non-linear Integrable Systems--Classical Theory and Quantum Theory: Proceedings of RIMS Symposium. Ed: M. Jimbo, T. Miwa. World Sci (US Distr: Heyden & Son), 1983, vi + 289 pp, $29. [ISBN: 9971-950-32-4] Proceedings of a May 1981 symposium held at Kyoto University. LAS

**Differential Geometry, S**(18), P.** Global Riemannian Geometry. Ed: T.J. Willmore, N.J. Hitchin. Ser. in Math. & Its Applic. Halsted Pr, 1984, 213 pp, $64.95. [ISBN: 0-470-20017-0] Twenty-two brief research-expository articles, forming an admirably well-written account of recent advances and open problems in Riemannian geometry. Sample: M.F. Atiyah, on Yang-Mills equations; S. Kobayashi, on Hermitian vector bundles; R. Penrose, on general relativity; S.T. Yau, on minimal surfaces. Definitely requires a good background in algebraic topology and differential geometry. Should bring a stimulating overview of the field to graduate students in quest of a dissertation topic. Relatively expensive. YN

**Differential Geometry, S(18).** Total Mean Curvature and Submanifolds of Finite Type. Bang-yen Chen. World Sci (US Distr: Heyden & Son), 1984, xi + 352 pp, $19 (P); $37. [ISBN: 9971-966-03-4; 9971-966-02-6] First, surveys Sard´s, Morse´s, de Rham´s, and Frobenius´ theorems, Hodge decomposition, etc., sometimes without purpose (says on p. 165 that tight immersions are important, but never defines nor uses them). Good review, but not an introductory course. Second, relates the spectrum of eigenvalues of the Laplacian on a compact Riemannian manifold to geometric notions such as volume, mean curvature, and minimal area. Interesting mathematics for the reader who doesn´t mind dangling grammar and is already quite familiar with differential geometry. YN

**Differential Geometry, T(18: 1), S**, P**.** Differentiable Manifolds: Forms, Currents, Harmonic Forms. Georges de Rham. Transl: F.R. Smith. Grund. der math. Wissenschaften, B. 266. Springer-Verlag, 1984, x + 167 pp, $28.50. [ISBN: 0-387-13463-8] Historical introduction by S.S. Chern. Excellent English translation of a classic that fills the gap between graduate textbooks and research papers in differential geometry. Discusses topics on "currents," some still difficult to find elsewhere, but underlying numerous recent developments. Direct and inverse images of currents (fundamental to double fibrations); homology of currents on manifolds (applies to minimal surfaces); double currents (yield smoothing kernels and fundamental solutions to partial differential equations); regularity of Green´s operator and harmonic currents. Culminates with de Rham´s proof of Hodge´s decomposition. An enriching supplementary reading. YN

**Algebraic Topology, S(18), P*.** Instantons and Four-Manifolds. Daniel S. Freed, Karen K. Uhlenbeck. Math. Sci. Res. Institute Pub., V. 1. Springer-Verlag, 1984, 232 pp, $15. [ISBN: 0-387-96036-8] The outcome of a seminar at the new Berkeley Institute which studied S. Donaldson´s proof that certain 4-dimensional manifolds, constructed by M. Freedman in his solution to the 4-dimensional Poincaré conjecture, are not smoothable. The proof of Donaldson´s result in topology uses the Yang-Mills equations of "ultra-modern" physics! Audience: the professional researcher. RB

**Algebraic Topology, S(18), P.** The Smith Conjecture. Ed: John W. Morgan, Hyman Bass. Academic Pr, 1984, xv + 243 pp, $49.50. [ISBN: 0-12-506980-4] Is the fixed point set of a periodic orientation preserving self-homeomorphism of $S^3$ always an unknotted circle? The 1978 affirmative answer for diffeomorphisms combined work of mathematicians in diverse areas within topology, geometry, algebra.

This volume includes presentations of the various pieces, as given by the principals at a 1979 Columbia symposium, plus an introduction and generalizations. RB

**Topology, S(17), P.** Fake Topological Hilbert Spaces and Characterizations of Dimension in Terms of Negligibility. J.J. Dijkstra. CWI Tract, V. 2. Math Centrum, 1984, iii + 109 pp, Dfl. 16,70 (P). [ISBN: 90-6196-268-4] An investigation in infinite-dimensional topology which looks at separable metrizable spaces that share many topological properties with $\ell^2$, but yet are not homeomorphic to it. List of references. No exercises. CEC

**Topology, P.** Lecture Notes in Mathematics-1060: Topology. Ed: L.D. Faddeev, A.A. Mal′cev. Springer-Verlag, 1984, vi + 389 pp, $18.50 (P). [ISBN: 0-387-13337-2] Proceedings of the International Topological Conference held in Leningrad, August 23-27, 1982. Contains papers on general topology, algebraic topology, and "applications" to related mathematical areas. JAS

**Operations Research, P.** Lecture Notes in Economics and Mathematical Systems-226: Selected Topics in Operations Research and Mathematical Economics. Ed: G. Hammer, D. Pallaschke. Springer-Verlag, 1984, ix + 478 pp, $27.50 (P). [ISBN: 0-387-12918-9] Proceedings of the Eighth Symposium on Operations Research held at the University of Karlsrühe from August 22-25, 1983. Includes thirty-seven papers on optimization theory, control theory, mathematical economics, game theory, graph theory, fixed point theory, statistics, measure theory, and applications. AO

**Operations Research, S(17-18), P.** Markovian Control Problems: Functional Equations and Algorithms. A. Federgruen. Math. Centre Tracts, No. 97. Math Centrum, 1983, vii + 212 pp, Dfl 31 (P). [ISBN: 90-6196-165-3] Author′s 1978 doctoral dissertation on Markovian control and optimization. Treats Markovian decision theory and many-person stochastic games, emphasizing successive approximation methods and algorithms. RM

**Operations Research, P, L.** A Model-Management Framework for Mathematical Programming. Kenneth H. Palmer, et al. Wiley, 1984, xix + 392 pp, $42.50. [ISBN: 0-471-80472-X] Describes a support system for mathematical programming applications developed at Exxon. All activities associated with linear programming (matrix generation, report writing, data management, model solution) are integrated into a single system. The system described is a model for the design of similar systems. AO

**Optimization, T(18), P.** Explicit Methods of Optimization. Jean-Pierre Aubin. Gauthier-Villars (US Distr: Heyden & Son), 1984, 287 pp, $44. [ISBN: 2-04-015575-9] Theoretical introduction to optimization with primary motivation applications to microeconomic theory. Discusses aspects of optimal control theory, quadratic and convex programming and applications to quadratic economic models. Appendix on the MODULECO package for macroeconomic models. RM

**Optimization, T(16-17), P.** Optimization: Theory and Applications, Second Edition. S.S. Rao. Wiley, 1984, xiv + 747 pp, $24.95. [ISBN: 0-470-27483-2] Unaltered reprint of the original 1978 edition (TR, May 1981). LAS

**Optimization, P.** Lecture Notes in Control and Information Sciences-59: System Modelling and Optimization. Ed: P. Thoft-Christensen. Springer-Verlag, 1984, ix + 934 pp, $49.50 (P). [ISBN: 0-387-13185-X] Ninety papers from the Eleventh IFIP Conference held in Copenhagen, Denmark, July 25-29, 1983. AO

**Optimization, P.** Progress in Combinatorial Optimization. Ed: William R. Pulleyblank. Academic Pr, 1984, xi + 374 pp, $43.50. [ISBN: 0-12-566780-9] Twenty-one papers presented at a conference held at the University of Waterloo during the summer of 1982. Represents a broad cross-section of current work in the field. AO

**Optimization, T(16-17: 1), S, L.** Introduction to Dynamic Programming. Leon and Mary W. Cooper. Intern. Ser. in Modern Appl. Math. & Comp. Sci. Pergamon Pr, 1981, ix + 289 pp, $19.50 (P). [ISBN: 0-08-025064-5] Relatively self-contained, readable introduction to dynamic programming. Nice balance among theory, methods, and applications, illustrated by interesting computational examples. Exercises, final chapter surveys some real applications. RM

**Probability, P.** Stochastic Functional Differential Equations. S-E A Mohammed. Research Notes in Math., V. 99. Pitman, 1984, vi + 245 pp, $21.95 (P). [ISBN: 0-273-08593-X] Existence of solutions and dependence on the initial process; Markov trajectories; the infinitesimal generator; regularity of the trajectory field; examples; further developments, problems and conjectures. Almost entirely new results. A first chapter provides background. DA

**Probability, P.** Probabilistic Analysis and Related Topics, Volume 3. Ed: A.T. Bharucha-Reid. Academic Pr, 1983, x + 260 pp, $47. [ISBN: 0-12-095603-9] Five survey articles on topics pertaining to analytic properties of random functions, the third volume in a series of occasional collections of papers on this topic. LAS

**Probability, P.** Lecture Notes in Mathematics-1064: Probability Measures on Groups VII. Ed: H. Heyer. Springer-Verlag, 1984, x + 588 pp, $27.70 (P). [ISBN: 0-387-13341-0] Proceedings of a conference held in Oberwolfach, West Germany, April 24-30, 1983. JAS

**Probability, T(15-17), S, P.** Statistical Analysis of Stationary Time Series. Ulf Grenander, Murray Rosenblatt. Chelsea, 1984, 308 pp, $17.95. [ISBN: 0-8284-0320-1] A corrected version of the 1957 edition. A "rigorous mathematical discussion" of the topic. FLW

**Probability, P.** Lecture Notes in Mathematics-1059: Séminaire de Probabilités XVIII: 1982/83. Ed: J. Azéma, M. Yor. Springer-Verlag, 1984, iv + 518 pp, $26 (P). [ISBN: 0-387-13332-1] Proceedings of the Paris Seminar for the academic year 1982-1983. JAS

**Statistics, T(15-17: 1, 2), S, L.** Statistical Inference. Vijay K. Rohatgi. Wiley, 1984, xiv + 940 pp, $44.95. [ISBN: 0-471-87126-5] Presupposes calculus. Introduces statistical ideas near the beginning and integrates parametric and non-parametric inferences. Considers the usual probability distributions and statistical tests, large sample theory, estimation, measures of association, contingency tables, and analysis of variance. No regression. Many exercises. FLW

**Statistics, T*(13-14: 1, 2), S*, L*.** Introductory Statistics for Business and Economics, Third Edition. Thomas H. and Ronald J. Wonnacott. Appl. Prob. & Stat. Wiley, 1984, xxii + 746 pp, $28.95. [ISBN: 0-471-09716-0] This new edition of this careful text (that presupposes only high school algebra) includes exploratory data analysis, the jackknife, path analysis, Box-Jenkins ARIMA forecasting, decision trees, and robust estimation as well as nonparametric texts, multiple regression, and Bayesian inference. (Second Edition, TR, March 1978.) FLW

**Statistics, S(14-18), L.** Statistical Modelling. Warren Gilchrist. Wiley, 1984, xv + 339 pp, $29.95. [ISBN: 0-471-90380-9] Attempts to "bring together and present in a structured fashion those statistical and related ideas that are needed for the construction of quantitative models." Includes chapters on identification, estimation, validation, and iteration. No exercises. FLW

**Statistics, P.** Proceedings of the Twenty-Ninth Conference on the Design of Experiments. US Army Research Office (PO Box 12211, Research Triangle Park, NC), 1983, xx + 356 pp, (P). Papers from an October 1983 conference held at the Uniformed Services University of Health Sciences in Bethesda, Maryland. LAS

**Statistics, T?(15-17: 1, 2), S.** Probability Theory and Mathematical Statistics for Engineers. V.S. Pugachev. Transl: I.V. Sinitsyna, P. Eykhoff. Pergamon Pr, 1984, xvii + 450 pp, $90. [ISBN: 0-08-029148-1] Probability, random variables and vectors, estimation of parameters and distributions, regression, analysis of variance, and brief treatments of some multivariate topics. No problem sets. Note the price! FLW

**Statistics, T(15-18: 1, 2), S, L.** Linear Statistical Models and Related Methods With Applications to Social Research. John Fox. Wiley, 1984, xx + 449 pp, $39.95. [ISBN: 0-471-09913-9] Presupposes calculus, matrix algebra, and an introduction to probability and statistics. Tries to blend theory and applications, with informal derivation or intuitive justifications given for most results. Examples are real data. Takes up linear regression, analysis of variance, linear-model problems, structural-equation models, and log and log-linear models. FLW

**Computer Literacy, S*(13-18), L*.** The RS-232 Solution. Joe Campbell. Sybex, 1984, xiv + 194 pp, $16.95 (P). [ISBN: 0-89588-140-3] A very thorough yet highly readable tutorial and reference for anybody who has experienced frustration when trying to connect computers or peripherals. It provides a description of the RS-232C "standard," instructions for making a few simple but very effective tools, and a number of case studies which are presented with remarkable good humor in the face of misinformation and poor design by manufacturers. Indeed, one of the aims of this book is to enable a run-of-the-mill user to get around the missing or wrong information that is usually provided. This book is a jewel. JAS

**Computer Programming, T(13: 1).** The Basics of BASIC. Alfredo C. Gomez. Holt, Rinehart & Winston, 1983, xii + 303 pp, $17.95 (P). [ISBN: 0-03-063069-X] Concise, no-nonsense, machine-independent, business-oriented introductory text. Includes chapters on strings, matrices, files. Discusses flowcharts and structured programming. Short exercise sets at ends of chapters. Mentions features of advanced versions in text. Quick reference guides in appendices for BASIC-PLUS, Applesoft II, TRS-80 Level II, Microsoft. KS

**Computer Programming, T(13: 1), S.** A Guide to Programming in Applesoft, Second Edition. Bruce Presley. D Van Nostrand Reinhold, 1984, x + 232 pp, $16.50 (P). [ISBN: 0-442-27249-9] Excellent text for beginning Applesoft programming. Many applications included but will need supplementary material if used for graphics. Written in clear, understandable language. TR

**Computer Programming, T(13-14: 1).** The COMAL Handbook, Second Edition (Now for the Commodore 64). Len Lindsay. Reston, 1984, xii + 467 pp, (P). [ISBN: 0-8359-0784-8] COMAL is a language designed to teach the elementary principles of computer programming. It is intended to be the "ideal" first language, by incorporating the best facilities of BASIC, Pascal, and LOGO. The language actually exists and can be run on a number of microcomputers, including the COMMODORE-64. MS

**Computer Programming, T(13: 1), S, L.** Graded Problems in Computer Science. Andrew D. McGettrick, Peter D. Smith. Addison-Wesley, 1983, vi + 314 pp, $14.95 (P). [ISBN: 0-201-13787-9] Language-independent text for first programming course. Covers programming constructs, good programming practice, some algorithm design strategies. Examples in readable pseudocode. Extensive, interesting problem sets range from easy to challenging. Majority of problems is numerical; some string processing. KS

**Computer Programming, T(13: 1).** Problem Solving and Structured Programming in BASIC. Elliot B. Koffman, Frank L. Friedman. Addison-Wesley, 1979, xvi + 444 pp, $21.95 (P). [ISBN: 0-201-03888-9] Text for introductory programming course using BASIC. Describes minimal BASIC, BASIC-PLUS and Dart-

mouth BASIC. Emphasizes designing algorithms using stepwise refinement. Most sample problems include data table, flow diagram and computer program. Uses if-then-else and while loops. Chapters on string processing, matrix commands, files. KS

**Computer Programming, T(13: 1).** This is BASIC: An Introduction to Computer Programming. Robert F. Sutherland. Macmillan, 1984, xv + 447 pp, $18.95 (P). [ISBN: 0-02-418370-9] Leisurely introduction to programming using minimal BASIC language (e.g., FOR-NEXT introduced in Chapter 12). Uses flowcharts. Some comments on good programming practice and debugging procedures. Numerous examples and exercises; most require no more than elementary algebra. Appendices on files and structured programming. KS

**Computer Programming, S(13).** Applied Basic for Microcomputers. Roy A. Boggs. Reston, 1984, x + 277 pp, $16.95 (P). [ISBN: 0-8359-0042-8] This book is a good resource for elementary file manipulation. Contains primarily business applications. Designed for use with Apples, TRS-80's, and IBM-PC's. Attempts to control for differences in each specific version of BASIC with varying degrees of success. TR

**Computer Programming, T(13-14).** Business Computing: A Structured Approach to BASIC on the PDP-11 and VAX-11. James F. Peters, III, Hamed M. Sallam. Reston, 1985, xviii + 684 pp, (P). [ISBN: 0-8359-0549-7] Focus is on business computing (especially accounting applications) and BASIC, with the intent of teaching both the language and how to use it to solve business problems. Many interesting, complete sample programs, with runs, but skimpy on exercises. RM

**Computer Programming, T(13: 1).** PASCAL. Charles H. Goldberg, Walter S. Brainerd, Jonathan L. Gross. Boyd & Fraser, 1984, xii + 465 pp, $22.95 (P). [ISBN: 0-87835-140-X] Text for first computer science course. Emphasizes top-down programming. Comments on style, testing, debugging, and common errors throughout. Introduces simple data structures; recursion, searching and sorting. Includes examples of simulation, graphics, computer-assisted instruction. No previous computing experience or college-level mathematics assumed. KS

**Software Systems, L.** Inside CP/M-86: A Guide for Users. David E. Cortesi. Holt, Rinehart & Winston, 1984, xv + 202 pp, $17.45 (P) [ISBN: 0-03-062656-0]; Inside CP/M Plus: A Guide for Users, 1984, xvi + 261 pp, $18.45 (P) [ISBN: 0-03-070671-8]; Inside Concurrent CP/M: A Guide for Users, 1984, xvii + 247 pp, $18.45 (P). [ISBN: 0-03-070669-6] These three books present a relaxed, even breezy, view of various branches of the CP/M operating system. The presentations are quite complete from the user's point of view. However, a person wanting deeper system information (e.g., disk directory structure) will find nothing. JAS

**Software Systems, P.** Lecture Notes in Computer Science-174: EUROSAM 84. Ed: John Fitch. Springer-Verlag, 1984, xi + 396 pp, $18 (P). [ISBN: 0-387-13350-X] Proceedings of the third of the international computer algebra conferences held in Europe every 5 years. The 37 papers span topics ranging from languages and algorithms for simplification, integration, factorization, number theory and computational group theory, to applications including differential equations, algebraic number computations and Groebner bases. RB

**Software Systems, L.** The CP/M Plus Handbook. Alan R. Miller. Sybex, 1984, xii + 248 pp, $15.95 (P). [ISBN: 0-89588-158-6] A straight-forward tutorial introduction and reference to CP/M-3, also known as CP/M Plus. The book is primarily aimed at the absolute novice though it does have a brief chapter on the directory structure and BDOS (system) calls. JAS

**Software Systems, P.** Approaches to Prototyping. Ed: R. Budde, et al. Springer-Verlag, 1984, xi + 458 pp, $23 (P). [ISBN: 0-387-13490-5] Edited papers from a working conference on "user-oriented software construction" held in Namur in West Germany in October 1983. JAS

**Computer Science, P.** Lecture Notes in Computer Science-169: Feedback Shift Registers. Christian Ronse. Springer-Verlag, 1984, 145 pp, $9 (P). [ISBN: 0-387-13330-5] Develops a general algebraic theory of feedback shift registers or "clocked switching circuits" over a finite field, using algebras of operators on sequences over the field, in which the identity, the shift, and the constants play a central role. Randomness properties are investigated. RM

**Computer Science, T(16-17: 1), P.** Distributed Computing. Ed: Fred B. Chambers, David A. Duce, Gillian P. Jones. Academic Pr, 1984, xii + 327 pp, $22.50. [ISBN: 0-12-167350-2] This text describes research activities in the area of distributed, or parallel, computing architectures. The new non-von Neumann architectures described in this text include the dataflow model, and high speed local networks. The text also describes research work in the area of programming languages for parallel systems (e.g., PROLOG), and distributed operating systems. MS

**Computer Science, P.** Dataflow Computation. A.P.W. Böhm. CWI Tract, No. 6. Math Centrum, 1983, iii + 208 pp, Dfl. 29,80 (P). [ISBN: 90-6196-272-2] A study of dataflow computation (a particular parallel processing architecture) focusing on issues of computational complexity and program correctness. AO

**Computer Science, T(15-17: 1, 2), S, L.** Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science. Martin D. Davis, Elaine J. Weyuker. Comp. Sci. & Appl. Math. Academic Pr, 1983, xix + 425 pp, $35. [ISBN: 0-12-206380-5] Readable text covering elementary recursion theory, formal languages and automata, propositional and predicate logic, computational complexity, classification of unsolvable problems. Assumes substantial programming experience. Flexi-

ble organization makes this suitable for many courses. Short exercise sets at end of most sections. KS

**Computer Science, P, L*.** Advances in Computers, Volume 23. Ed: Marshall C. Yovits. Academic Pr, 1984, xiii + 377 pp, $48. [ISBN: 0-12-012123-9] Seven expository papers on supercomputers and VLSI, information and computation, videogames, decision support systems, digital control systems, information privacy, and parallel sorting algorithms. LAS

**Computer Science, P.** Lecture Notes in Computer Science-176: Mathematical Foundations of Computer Science 1984. Ed: M.P. Chytil, V. Koubek. Springer-Verlag, 1984, xi + 581 pp, $25 (P). [ISBN: 0-387-13372-0] Papers contributed to the 11th Symposium on Mathematical Foundations of Computer Science in Prague, September 1984. Emphasis on complexity and computability questions related to logic, formal languages, and automata. RM

**Computer Science, P.** Information Technology and the Computer Network. Ed: Kenneth G. Beauchamp. NATO ASI Ser. F: Computer & Systems Sci. Springer-Verlag, 1984, viii + 271 pp, $34.50. [ISBN: 0-387-12883-2] Collection of papers presented at the NATO Advanced Study Institute on Information Technology and the Computer Network, 1983. Intent is to provide a study of the current achievements in network technology. JRG

**Control Theory, S(16-18), P*.** Control and Dynamic Systems, Volume 19: Nonlinear and Kalman Filtering Techniques, Part 1 of 3. Ed: C.T. Leondes. Academic Pr, 1983, xvii + 373 pp, $37.50. [ISBN: 0-12-012719-9] Five papers on nonlinear dynamic systems, applications of Bayesian estimation techniques, and computational techniques. AO

**Control Theory, P.** Control and Dynamic Systems: Advances in Theory and Applications, V. 20: Nonlinear and Kalman Filtering Techniques, Part 2 of 3. Ed: C.T. Leondes. Academic Pr, 1983, xix + 410 pp, $44. [ISBN: 0-12-012720-2] Seven papers make up the second volume in a trilogy on nonlinear and Kalman filters. LAS

**Control Theory, T(17: 1), P.** Oscillation Theory of Optimal Processes. George M. Smirnov. Wiley, 1984, x + 154 pp, $29.95. [ISBN: 0-471-89399-4] The set of all control functions that satisfy the Pontryagin maximum principle is classified on the basis of an analysis of oscillatory properties. Growing out of results first obtained in 1971-1977, this book includes additional material proved in 1980-1983. AWR

**Control Theory, P.** Lie Groups: History, Frontiers and Applications, Series B: Systems Information and Control, Volume II. Ed: L.R. Hunt, C.F. Martin. Math Sci Pr, 1984, vi + 450 pp, $50. [ISBN: 0-915692-37-6] Proceedings of the Berkeley-Ames Conference on nonlinear problems in control and fluid dynamics held in Berkeley, May 1983. JAS

**Applications, P*.** Ocean Tides: Mathematical Models and Numerical Experiments. G.I. Marchuk, B.A. Kagan. Transl: E.V. Blinova, L. Ya. Yusina. Pergamon Pr, 1984, xii + 292 pp, $65. [ISBN: 0-08-026236-8] Account of recent (1960's and 1970's) work by and views of Russian authorities on tides. Much material not hitherto available in English. For researchers in oceanography and solid earth geophysics. JK

**Applications, P.** Lecture Notes in Computer Science-168: Methods and Tools for Computer Integrated Manufacturing. Ed: U. Rembold, R. Dillmann. Springer-Verlag, 1984, xvi + 528 pp, $20.80 (P). [ISBN: 0-387-12926-X] Lectures from a September 1983 advanced course held at the University of Karlsruhe. LAS

**Applications (Astronomy), L.** Cosmic Discovery: The Search, Scope, and Heritage of Astronomy. Martin Harwit. MIT Pr, 1984, xi + 334 pp, $9.95 (P). [ISBN: 0-262-58068-3] With respect to the discipline of astronomy, the author addresses the questions: "What do we know now?", "What was the process of discovery?", and "How should we plan for further discovery?" Interesting and lucid, but rather sophisticated. JRG

**Applications (Biology), P.** Lecture Notes in Biomathematics-54: Mathematical Ecology. Ed: S.A. Levin, T.G. Hallam. Springer-Verlag, 1984, xii + 513 pp, $24.50 (P). [ISBN: 0-387-12919-7] Proceedings of the research part of a December 1982 conference held at the International Center for Theoretical Physics in Miramare-Trieste, Italy. Papers deal with autoecology, population biology, ecosystem theory, diffusion models, and applications to fisheries and epidemiology. (Expository lectures from the first part of the conference are scheduled to be published separately as a textbook.) LAS

**Applications (Chemistry), P.** Lecture Notes in Mathematics-1063: Orienting Polymers. Ed: J.L. Ericksen. Springer-Verlag, 1984, vii + 166 pp, $9.50 (P). [ISBN: 0-387-13340-2] Proceedings of a workshop held March 21-26, 1983 at the University of Minnesota. These papers provide an introduction to the problems of interest and a survey what is known about them. AO

**Applications (Economics), P.** Operations Research and Economic Theory: Essays in Honor of Martin J. Beckmann. Ed: H. Hauptmann, W. Krelle, K.C. Mosler. Springer-Verlag, 1984, xi + 378 pp, $28.50. [ISBN: 0-387-13652-5] A hard-to-classify range of papers running from air pollution reductions in Munich to a discussion of whether the ellipsoid method can ever be an efficient way to solve linear programming methods. AWR

**Applications (Economics), P.** Lecture Notes in Economics and Mathematical Systems-229: Interactive Decision Analysis. Ed: M. Grauer, A.P. Wierzbicki. Springer-Verlag, 1984, viii + 268 pp, $15.50 (P). [ISBN: 0-387-13354-2] Proceedings of an international workshop on interactive decision analysis and interpretative computer intelligence held in Laxenburg, Austria, September 20-23, 1983. JAS

**Applications (Economics), P*.** Lecture Notes in Economics and Mathematical Systems-230: Macro-Economic Planning with Conflicting Goals. Ed: M. Despontin, P. Nijkamp, J. Spronk. Springer-Verlag, 1984, vi + 297 pp, $17.50 (P). [ISBN: 0-387-13367-4] The principal novelty lies in quite new approaches to solve macro-economic problems mathematically, rather than in new mathematics. Utilizes mainly game theory, mathematical programming, and fuzzy sets in order to rank the policy makers' preferences, resolve conflicts between competing goals, and compare outcomes. Reports genuine case studies, for instance interactions between European countries, the Tasmanian economy, and water treatment on San Francisco Bay. YN

**Applications (Economics), P.** Lecture Notes in Economics and Mathematical Systems-227: Risk and Capital. Ed: Günter Bamberg, Klaus Spremann. Springer-Verlag, 1984, vi + 306 pp, $18 (P). [ISBN: 0-387-12923-5] Proceedings of the second international workshop on risk and capital held at the University of Ulm in June 1983. LAS

**Applications (Economics), P.** Economics Consequences of Population Change in Industrialized Countries. Ed: Gunter Steinmann. Stud. in Contemp. Econ., V. 8. Springer-Verlag, 1984, x + 415 pp, $23.50 (P). [ISBN: 0-387-13513-8] Papers from a June 1983 conference held at the University of Paderborn assessing with mathematical and economic models the impact of low fertility rates. All papers are in English, by German authors. LAS

**Applications (Engineering), P.** Problems of Randomness in Communication Engineering. Ed: Kenneth W. Cattermole, John J. O'Reilly. Math. Topics in Telecommunications, V. 2. Wiley, 1984, 342 pp, $32.50. [ISBN: 0-471-80763-X] A tutorial introduction to the mathematics of stochastic systems and processes together with applications of these ideas in communications theory. AO

**Applications (Engineering), T(17-18: 1), P.** Elements of Digital Satellite Communication, Volume I. William W. Wu. Computer Sci Pr, 1984, xv + 607 pp, $44.95. [ISBN: 0-914894-39-0] An introduction to the design, planning, and operation of digital satellite communication systems. AO

**Applications (Engineering), T(16-17: 1), S.** The Finite Element Method Displayed. Gouri Dhatt, Gilbert Touzot. Wiley, 1984, xv + 509 pp, $39.95. [ISBN: 0-471-90110-5] Very attractive text for engineering students. Many, many examples. FORTRAN programs. Main topics: approximation of the unknowns, integral formulation, matrix formulation, numerical methods, programming techniques. Translated from French by Gilles Cantin. DA

**Applications (Engineering), P.** Numerical Methods in Coupled Systems. Ed: R.W. Lewis, P. Bettess, E. Hinton. Ser. in Num. Methods in Engin. Wiley, 1984, xiv + 618 pp, $74.95. [ISBN: 0-471-90122-9] This volume grew out of the International Conference on Numerical Methods for Coupled Problems, Swansea, Wales, held September 1, 1981 in honor of O.C. Zienkiewicz in his 60th year. JAS

**Applications (Engineering), T(18: 1), P.** Variational Inequalities and Flow in Porous Media. M. Chipot. Appl. Math. Sci., No. 52. Springer-Verlag, 1984, vii + 118 pp, $16 (P). [ISBN: 0-387-96002-3] Contents of one-semester graduate course given in 1981-82 at Brown. Two-chapter introduction to the variational inequality and obstacle problems. Chapter 3 is on regularity theory. Chapter 4 is on the dam problem. No exercises or problems. References. JK

**Applications (Management), P.** Lecture Notes in Economics and Mathematical Systems-228: Nonlinear Models of Fluctuating Growth. Ed: R.M. Goodwin, M. Krüger, A. Vercelli. Springer-Verlag, 1984, xvii + 277 pp, $17.50 (P). [ISBN: 0-387-13349-6] Revised papers from an international symposium held in Siena, Italy, March 24-27, 1983. JAS

**Applications (Management), P.** Mathematical Methods in Management. Geoffrey Gregory. Handbook of Applicable Math. Wiley, 1983, xii + 154 pp, $31.95. [ISBN: 0-471-90131-8] A volume in a series addressed to professional people who discover that their career demands more mathematics than their academic training provided. It is designed to accompany the six core volumes covering basic mathematical concepts. AWR

**Applications (Medicine), S(15-16), L.** Mathematical Methods in Medicine. Richard Bellman. World Scientific (US Dist: Heyden & Sons), 1983, xiv + 252 pp, $18 (P); $33. [ISBN: 9971-950-21-9; 9971-950-20-0] Mathematical modelling of physiological phenomena and the attendant analytic difficulties, using methods from differential equations, control theory and dynamic programming. Bibliography. JRG

**Applications (Medicine), P.** Applications of Optimal Control Theory in Biomedicine. George W. Swan. Pure & Appl. Math., No. 81. Dekker, 1984, xi + 285 pp, $55. [ISBN: 0-8247-7192-3] Two chapters on mathematical background, followed by self-contained chapters on diabetes, thyroid disease, administration of drugs, cancer therapy, and the circulatory systems. AWR

**Applications (Physics), P.** Lecture Notes in Mathematics-1048: Kinetic Theories and the Boltzmann Equation. Ed: C. Cercignani. Springer-Verlag, 1984, vii + 243 pp, $13 (P). [ISBN: 0-387-12899-9] Lectures from the Centro Internazionale Matematico Estivo held in Montecatini, Italy, June 10-18, 1981. JAS

**Applications (Physics), P.** Lecture Notes in Mathematics-1055: Quantum Probability and Applications to the Quantum Theory of Irreversible Processes. Ed: L. Accardi, A. Frigerio, V. Gorini. Springer-Verlag, 1984, vi + 411 pp, $19 (P). [ISBN: 0-387-12915-4] Proceedings of a September 1982 workshop held at Villa Mondragone, Italy. LAS

**Applications (Physics), S(18), P.** Advances in Nonlinear Waves, V. I. Ed: Lokenath Debnath. Research Notes in Math., V. 95. Pitman, 1984, 328 pp, $25.95 (P). [ISBN: 0-273-08609-X] A diverse collection of highly technical papers. MU

**Applications (Physics), S(16), P, L.** Concepts of Particle Physics, V. I. Kurt Gottfried, Victor F. Weisskopf. Oxford U Pr, 1984, xv + 189 pp, $22.50. [ISBN: 0-19-503392-2] This, the first of two volumes, is a descriptive survey of modern particle physics, virtually devoid of mathematics. MU

**Applications (Physics), P.** Monopoles in Quantum Field Theory. Ed: N.S. Craigie, P. Goddard, W. Nahm. World Scientific (US Dist: Heyden & Son), 1982, xxi + 440 pp, $21 (P); $49. [ISBN: 9971-950-29-4; 9971-950-28-6] Proceedings of the Monopole Meeting held in Trieste, Italy, December 1981. JAS

**Applications (Physics), P.** Lecture Notes in Physics-195: Trends and Applications of Pure Mathematics to Mechanics. Ed: P.G. Ciarlet, M. Roseau. Springer-Verlag, 1984, 422 pp, $20.50 (P). [ISBN: 0-387-12916-2] Proceedings of a symposium (fifth in a series) held in Palaiseau, France from November 28 to December 2, 1983. LAS

**Applications (Physics), T(16-17: 1, 2), S, L.** Mathematical Methods with Applications to Problems in the Physical Sciences. Ted Clay Bradbury. Wiley, 1984, ix + 702 pp, $37.95. [ISBN: 0-471-88639-4] Primarily for majors in physics. Prerequisites include calculus, differential equations and basic physics. Classical mechanics, electromagnetism and quantum mechanics are developed along with the mathematics. Emphasis on linear algebra. Good chapters on vectors and tensors and on curvilinear coordinates. Nice treatment of computer-oriented numerical methods. Wide coverage is sacrificed for completeness. Numerous problems--routine and otherwise. JK

**Applications (Physics), P*.** Applications of the Monte Carlo Method in Statistical Physics. Ed: K. Binder. Topics in Current Physics, V. 36. Springer-Verlag, 1984, xiv + 311 pp, $32. [ISBN: 0-387-12764-X] Nine papers that survey recent developments in the use of Monte Carlo techniques in statistical physics. AO

**Applications (Physics), P*, L?** Sound Propagation in Stratified Fluids. Calvin H. Wilcox. Appl. Math. Sci., No. 50. Springer-Verlag, 1984, ix + 198 pp, $19.80 (P). [ISBN: 0-387-90986-9] Presents a theory of the propagation of transient sound waves in fluids whose densities and sound speeds are functions of the depth. Highly mathematical with unusual clarity. The numerous proofs are carefully presented. MU

**Applications (Psychology), P, L.** Computation and Cognition: Toward a Foundation for Cognitive Science. Zenon W. Pylyshyn. MIT Pr, 1984, xxiii + 292 pp, $25. [ISBN: 0-262-16098-6] A systematic analysis of the science of "informavores"--consumers and processors of information--that is to the foundations of cognitive science "what a hole in the ground is to the foundation of a house." Pylyshyn--a psychologist and computer scientist--argues that cognition is a species of computation, that the semantic contents of mental states are encoded in the same general way as computer representations are encoded. LAS

**Applications (Psychology), T(16), S, P.** Mathematical Models of Attitude Change, Volume 1: Change in Single Attitudes and Cognitive Structure. John E. Hunter, Jeffrey E. Danes, Stanley H. Cohen. Human Commun. Res. Ser. Academic Pr, 1984, xv + 339 pp, $59. [ISBN: 0-12-361901-7] Translates existing verbal models of communication and information processing into mathematical models; suggests alternate models, and develops new theories of changes in attitude toward the source of information. Reports statistical analyses of field studies to fit the models to the data, often a difficult task. Assumes only a familiarity with calculus and matrix algebra, but then uses regressions to compute multivariate correlation matrices, and swiftly shoves the particulars of a linear system of difference equations into a quoted computer program (Euler's method?). A book for gifted students in psychology, communication, marketing, and related fields. YN

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## ON GELBAUM'S ALGORITHM FOR COMPUTING THE MINIMAL POLYNOMIAL OF A MATRIX

RONALD J. LEACH
*Department of Mathematics, Howard University, Washington, DC 20059*

B. Gelbaum [1] recently gave an algorithm for computing the minimal polynomial of an $n \times n$ matrix $A$ by embedding the matrix in row-major form in $\mathbb{C}^{n^2}$ and then applying the Gram-Schmidt process. The purpose of this note is to compare the complexity of his algorithm with existing algorithms.

We use Gelbaum's notation as follows. Let $\phi(A)$ denote the embedding of $A$ in $\mathbb{C}^{n^2}$, by spreading out the $n$-rows into one long row, and let

$$O_0 = \phi(I)/\sqrt{n},$$

$$X_{p+1} = \phi(A^{p+1}) - \sum_{k=0}^{p} \langle \phi(A^{p+1}), O_k \rangle O_k,$$

$$O_{p+1} = X_{p+1}/\|X_{p+1}\|,$$

be the vectors in $\mathbb{C}^{n^2}$ determined by the Gram-Schmidt process. The minimal polynomial $A_0 + A_1\lambda + \cdots + A_m\lambda^m$ has $A_k = \langle \phi(A^m), O_k \rangle$. The degree $m$ is the first integer for which $X_m = 0$. To compute the inner product of two $n^2$-vectors we may use $n^2$ multiplications and $(n^2 - 1)$ additions. Note that all operations are on complex numbers. As is common in arithmetic complexity theory (see for example S. Winograd [3]), we ignore the number of additions. To compute $\langle \phi(A^{p+1}), O_k \rangle O_k$ we must again compute $n^2$ multiplications of numbers. Computation of the norm of an $n^2$-vector requires $n^2$ additions and the taking of a square root, which will be done by Newton's method and whose complexity is independent of $n$. Finally there is a division step. Hence the multiplicative complexity of the minimal polynomial is bounded above by:

$$\begin{aligned}
&\left[ n^2(\text{multiplications for each inner product}) \right] \\
&+ \left[ n^2(\text{multiplications of the entries } O_k \text{ by the inner product}) \right] \\
&\times \left[ n(\text{number of terms in worst case}) \right] \\
&\times \left[ 2n^2 \begin{pmatrix} \text{number of squarings to compute norm } + \\ \text{number of divisions of } n^2\text{-vector by norm} \end{pmatrix} \right] \\
&\times C(\text{constant for computing square roots}) \\
&= O(n^5)
\end{aligned}$$

with a constant essentially equal to twice that for computing square roots.

There are a number of methods for reducing the number of multiplication steps in computing the inner product. The best results, due to Winograd, reduce the number of multiplications to approximately $3n^2/4$ at the expense of increasing the number of additions and complexity of the program. Since we are only considering orders of magnitude, these refinements will not be used.

The traditional method [2, p. 286] requires computing the determinant of $D = \lambda I - A$ and dividing by the greatest common divisor of the $(n - 1)$ minors of $D$. The classical computation of a determinant by row reduction requires $n$ comparisons in the first column with $(n - 1)$

multiplications of ($n$ column) rows by scalars for a total of $n^2 - n$ multiplications. After the $n^2 - n$ subtractions, we consider the second column which has $(n-1)$ multiplications of $(n-2$ column) rows by scalars or $(n-1)(n-2)$. (It is sufficient to transform $A$ to triangular form.)

In general we have

$$n(n-1)+(n-1)(n-2)+\cdots+2(1)$$
$$=\tfrac{1}{3}n^3 + \text{lower order terms}$$

for the number of multiplications to get $D$ in triangular form. There will then be $n$ multiplications (of diagonal entries) to compute $\det(D)$, requiring $O(n^4)$ multiplications.

As before, we do not consider the more sophisticated algorithms for computing determinants. We remark that there is an algorithm of V. Strassen that uses divide-and-conquer to reduce the number of multiplications but with a similar order of magnitude.

A similar situation holds for the computation of each of the $(n-1)$-minors of $D$. Since there are $n-1$ such minors, the multiplicative complexity is $(n-1)O(n^4) = O(n^5)$ with constant no more than $1/3$. We compute the gcd of the polynomial which are the $(n-1)$-minors of $D$ using the Euclidean algorithm on the first two, then on their gcd and the third, etc. An upper bound for the number of multiplication/division steps when dividing two polynomials of degree $n-1$ is $n$ divisions and $O(n^2)$ multiplications. Since there are $n-1$ polynomials, there are $O(n^4)$ such computations.

Thus the classical method requires

$$O(n^4)+O(n^5)+O(n^4)=O(n^5)$$

multiplication/division steps with the coefficient of $n^5$ at most 1. This is theoretically of the same order as Gelbaum's algorithm, but with a smaller constant. However, the analysis of Gelbaum's algorithm assumed a worst case of degree $n$. If for example the minimal polynomial is linear, then Gelbaum's algorithm takes $O(n^4)$ steps while the classical one requires $O(n^5)$ steps. Exactly the opposite situation occurs if the minimal polynomial is the characteristic polynomial.

There is one final factor to be considered. With the rapid increase in computing power and decrease in cost, a major factor in cost is programmer time. We invite the interested reader to design and program an algorithm for the traditional procedure.

## References


1. B. Gelbaum, An algorithm for the minimal polynomial of a matrix, this MONTHLY, 90 (1983) 43–44.
2. J. Moore, Elements of Linear Algebra and Matrix Theory, McGraw-Hill, New York, 1968.
3. S. Winograd, Arithmetic Complexity of Computations, CBMS-NSF Regional Conference Series 33, SIAM, Philadelphia, 1980.


# CONTINUOUS FUNCTIONS WITH A DENSE SET OF PROPER LOCAL MAXIMA


VLADIMIR DROBOT
*Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053*

MICHAL MORAYNE
*Instytut Matematyczny, Uniwersytet Wroclawski, Wroclaw, Poland*


We say that a real valued function $f\colon [0,1] \to R$ has a proper local maximum at $x \in [0,1]$ if there exists $\varepsilon > 0$ such that $f(y) < f(x)$ for all $y \in [0,1]$ satisfying $0 < |y-x| < \varepsilon$.

The sets of such proper local maxima have attracted some attention; see the references. For any function $f$, the set of its proper local maxima is known to be at most countable. A strong converse is due to Z. Zalcwasser [**8**] and states that for any two disjoint countable sets $A, B \subseteq [0,1]$, there exists a differentiable function $f\colon [0,1] \to R$ such that $A$ and $B$ are,

respectively, exactly the sets of its proper maxima and minima. For a new proof of this theorem see [3]. In a recent paper [5] in this MONTHLY, E. E. Posey and J. E. Vaughan constructed an explicit example of a continuous function which assumes a proper local maximum at each point of a dense set. The construction is very reminiscent of a construction of a continuous nowhere differentiable function. This analogy goes further. As in the case of nowhere differentiable functions (see [4], p. 45), the existence of a continuous function with a dense set of proper local maxima can also be proved by the Baire category method. More precisely, let $C[0,1]$ denote the space of all continuous real valued functions defined on $[0,1]$ with the metric

$$d(f,g) = \sup\{|f(x) - g(x)|: x \in [0,1]\}$$

(this is a complete metric space) and let $D$ be the subset of $C[0,1]$ consisting of all those functions whose set of proper local maxima is dense in $[0,1]$.

Recall that a subset of a metric space is said to be of first category if it is the union of a countable collection of sets each of which has a closure with empty interior. A subset of a metric space is called residual if its complement is of first category. The Baire Category Theorem states that in a complete metric space the empty set is not residual. In this note we give a straightforward proof that $D$ is residual in $C[0,1]$, thus showing that $D$ is nonempty. Our proof is certainly simple enough to be used as an exercise in any course which discusses the Baire Category Theorem. The result itself is not really new; for example it can be derived from Theorem 2.1, p. 212 and Lemma 3.1, p. 215 of [1]. We will also show that $D$ is a Borel subset of $C[0,1]$.

We introduce some notations. The letter $I$ will always denote a closed subinterval of $[0,1]$, $I^0$ will denote the interior of $I$, and $|I|$ its length. A rational interval is an interval with rational endpoints. Given any interval $I \subset [0,1]$ define

$$A(I) = \{f \in C[0,1]: \text{there is } x \in I^0 \text{ such that } f(x) > f(y) \text{ for all } y \in I \setminus \{x\}\}.$$

Let $J \subset I^0$ and put

$$B(I,J) = \{f \in C[0,1]: \sup\{f(x): x \in J\} > \sup\{f(x): x \in I \setminus J^0\}\}.$$

THEOREM. *The set $D$ is a residual Borel subset of $C[0,1]$.*

*Proof.* We first show that for every interval $I$, the set $A(I)$ is dense in $C[0,1]$. Let $f \in C[0,1]$ and let $\varepsilon > 0$. Let $z = \sup\{f(y): y \in I\}$. Choose a nontrivial open interval $J \subset I$ such that $f(x) > z - \varepsilon$ for every $x \in J$. Now define $g \in C[0,1]$ as follows: $g$ equals $z + \varepsilon$ at the midpoint of $J$, $g$ equals $f$ at both endpoints of $J$; on $J$ the graph of $g$ consists of two straight line segments connecting the three points already defined; on $I \setminus J$, $g$ equals $f$. Now it is clear that $g \in A(I)$ and $d(f,g) \leqslant 2\varepsilon$. Thus $A(I)$ is dense in $C[0,1]$, as desired.

Next we show that $A(I)$ is a $G_\delta$ set. For each $J \subset I^0 \subset [0,1]$, the set $B(I,J)$ is open. Indeed if $f \in B(I,J)$, then the open ball with center $f$ and radius

$$\tfrac{1}{3}\left(\sup\{f(x): x \in J\} - \sup\{f(x): x \in I \setminus J^0\}\right)$$

is entirely contained in $B(I,J)$. We shall show for each closed interval $I \subseteq [0,1]$ that

$$A(I) = \bigcap_{i=1}^{\infty}\left\{\cup B(I,J): J \text{ is a closed interval}, \quad J \subset I^0, \quad |J| < \frac{1}{i}\right\}. \tag{1}$$

Suppose then that $f$ belongs to the right hand side of (1). This means that for each $i = 1,2,3,\ldots$, there is a closed interval $J_i$ such that $f \in B(I,J_i)$ and $|J_i| < 1/i$. From the definition of $B(I,J)$ it is clear that $\bigcap_{i=1}^{\infty} J_i$ contains all points where $f$ attains its maximum on $I$. Since $|J_i| \to 0$, $\bigcap_{i=1}^{\infty} J_i$ consists of a single point $x$. Because each $J_i \subset I^0$, we see that $x \in I^0$ and that $f$ attains its maximum on $I$ at $x$ and only at $x$. Thus $f \in A(I)$. The opposite inclusion is obvious.

Thus the set $A(I)$ is a dense $G_\delta$ subset of $C[0,1]$, so it is residual. Therefore the set

$$X = \cap\{A(I): I \subset [0,1],\ I \text{ rational}\}$$

is also residual. It is also quite clear that $X \subset D$ so that $D$ is residual too.

To see that $D$ is a Borel set it is enough to notice that

$$D = \bigcap_{T \subset [0,1],\, T \text{ rational}} \cup \{ A(I) : I \subset T,\, I \text{ rational} \}.$$

That finishes the proof.

It is perhaps worth mentioning that it we define a set $E$ to consist of all functions with a dense set of proper minima, then $E$ is also residual by a similar argument. Thus the set $D \cap E$ consisting of all functions which have their sets of proper minima and maxima dense in $[0,1]$ is again residual and thus nonempty. To construct a specific example of a function $f$ with such property along the lines of [**5**] would probably be quite complicated.

We are indebted to Professor K. M. Garg for helpful discussions concerning this paper.

**References**

**1.** A. M. Bruckner, Differentiation of Real Functions, Lecture Notes in Mathematics 659, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

**2.** A. Denjoy, Mémoire sur les nombres dérivés des fonctions continues, J. Math. Pures Appl., 70, 1 (1915) 105–240.

**3.** V. Kelar, On strict local extrema of differentiable functions, Real Analysis Exchange, 6, 2 (1980–1981) 242–244.

**4.** J. C. Oxtoby, Measure and Category, Graduate Texts in Mathematics 2, Springer-Verlag, New York-Heidelberg-Berlin, 1980.

**5.** E. E. Posey and J. E. Vaughan, Functions with a proper local maximum in each interval, this MONTHLY, 90 (1983) 281–282.

**6.** A. Schoenflies, Die Entwickelung der Lehre von dem Punktmannigflatigkeiten, Jahresbericht Deutschen Mathematiker-Vereinigung 8, Leipzig, 1900.

**7.** W. Sierpinski, Sur l'ensemble des points angulaires d'une courbe $y = f(x)$, Bull. Acad. Sci., Cracovie (1912) 850–855.

**8.** Z. Zalcwasser, Sur les fonctions de Kopcke, Prace Mat. Fiz., 35 (1927–28) 57–99.

## THE DIAMETER OF A GRAPH AND ITS COMPLEMENT

FRANK HARARY
*Department of Mathematics, University of Michigan, Ann Arbor, MI 48109*

ROBERT W. ROBINSON
*Department of Mathematics, Southern Illinois University, Carbondale, IL 62901*

**Dedicated to Gerhard Ringel and Horst Sachs**

The most active mathematician in the world [**1**] is well known for his personalized, picturesque terminology for kinship relations. He has also proposed what might be called "The Divine Book of Proofs", thereby extending significantly the scope of the Platonist philosophy of mathematics. In this apocryphal book all theorems which can ever be discovered are included, together with the best possible, most elegant proof of each. We offer a candidate for such a proof of the result, independently discovered by Ringel [**3**] and Sachs [**4**], that every nontrivial self-complementary graph has diameter 2 or 3. This will follow as an immediate corollary of an easy theorem on the diameter of a graph and its complement. We follow the notation of [**2**], where definitions not included here may be found.

A few concepts about graphs are needed. The *complete graph* $K_p$ has $p$ points with every two points adjacent (joined by a line). Let $G$ be a graph having the same points as $K_p$. By definition, the *complement* $\overline{G}$ of graph $G$ also has this point set and contains just those lines of $K_p$ which are

not in $G$. Then $G$ is *self-complementary* when $\bar{G}$ and $G$ are isomorphic. The *distance* $d_G(u,v)$ is the minimum length of a $u$-$v$ path in $G$. The *diameter* $d(G)$ is the maximum distance which occurs in $G$, with the convention that $d(G) = \infty$ when $G$ is disconnected.

THEOREM. *If* $d(G) \geqslant 3$, *then* $d(\bar{G}) \leqslant 3$.

*Proof*. As the diameter $d(G) \geqslant 3$, there are two points $u$ and $v$ of $G$ at distance $d_G(u,v) \geqslant 3$. Then in the complement, $d_{\bar{G}}(u,v) = 1$ at once. Now let $x$ and $y$ be any two points of $G$. By the hypothesis $u$ and $v$ cannot both be adjacent in $G$ to $x$, so that $d_{\bar{G}}(u,x) = 1$ or $d_{\bar{G}}(v,x) = 1$, say the former (without loss of generality). Similarly $d_{\bar{G}}(u,y) = 1$ or $d_{\bar{G}}(v,y) = 1$. Thus in $\bar{G}$ there is a path $xuy$ or $xuvy$, and $d_{\bar{G}}(x,y) \leqslant 3$ as required.

COROLLARY. *Every nontrivial self-complementary graph G has diameter* $d(G) = 2$ *or* 3.

*Proof*. Every nontrivial graph $G$ has diameter $d \geqslant 1$, with equality only if $G$ is complete. If $G$ is also self-complementary, then it cannot be complete, so $d(\bar{G}) \geqslant 2$. As the diameter is a graphical invariant, $d(\bar{G}) = d(G)$. Now there are two possibilities: either $d(G) = 2$ or $d(G) \geqslant 3$. In the latter case, the theorem guarantees that $d(G) \leqslant 3$ and hence that $d(G) = 3$.

Note that the two self-complementary graphs with five points, the pentagon $C_5$ and the graph shaped like the letter A, have diameter 2 and 3, respectively.

**References**

1. P. Erdős, Extremal problems in graph theory, in "A Seminar on Graph Theory," Holt, New York, 1967, pp. 54–59.
2. F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
3. G. Ringel, Selbstkomplementäre Graphen, Arch. Math., 14 (1963) 354–358.
4. H. Sachs, Über selbstkomplementäre Graphen, Publ. Math. Debrecen, 9 (1962) 270–288.

## THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

### SIMPLIFICATION OF SOME CONTOUR INTEGRATIONS

MARY L. BOAS
*Department of Physics, DePaul University, Chicago, IL 60614*

R. P. BOAS
*Department of Mathematics, Northwestern University, Evanston, IL 60201*

In introductory courses on complex analysis we learn to evaluate interesting real integrals by integrating around suitable contours. Common examples are $\int_0^\infty x^{-1}\sin x\,dx$, for which we integrate $z^{-1}e^{iz}$ around a semicircle (center at 0, diameter along the real axis); or the Fresnel integrals $\int_0^\infty \sin x^2\,dx$ and $\int_0^\infty \cos x^2\,dx$, for which we integrate $z^{-1/2}e^{iz}$ around a quarter-circle. In such problems we end up having to show that

$$\lim_{R\to\infty} R^\lambda \int_0^{\pi/2} e^{-R\sin\theta}\,d\theta = 0,$$

where $\lambda = 1/2$ for the Fresnel integrals, and $\lambda = 0$ for integrals of the form $\int_{-\infty}^{\infty} e^{ix}R(x)\,dx$,

where $R$ is a rational function with the degree of the denominator greater than the degree of the numerator. The integral involved in the limit is not elementary, and has to be estimated by some device before we can say that its limit is 0. The naive argument that $R^\lambda e^{-R\sin\theta} \to 0$ for each $\theta$ except 0 is inadequate, since it would purport to show equally well that $R\int_0^{\pi/2} e^{-R\sin\theta}\, d\theta \to 0$, which is false. Indeed, if this were true it would follow that $\int e^{iz}\, dz$ along $|z| = R$ from $\theta = 0$ to $\pi/2$ tends to zero as $R \to \infty$. This would lead to $\int_0^\infty e^{ix}\, dx = \int_0^\infty e^{-y}\, dy$, although the first integral does not converge.

The usual device is to apply the inequality $\sin\theta \geqslant 2\theta/\pi$ in order to replace the integral by one that can be evaluated explicitly. Students often find this step difficult because they do not know the inequality, and one has to digress in order to prove it.

An easier device is to break the integral into $\int_0^{\pi/3} + \int_{\pi/3}^{\pi/2}$. The second integral causes no difficulty; in the first, since $2\cos\theta \geqslant 1$ for $0 \leqslant \theta \leqslant \pi/3$, we have

$$R^\lambda \int_0^{\pi/3} e^{-R\sin\theta}\, d\theta \leqslant 2R^\lambda \int_0^{\pi/3} e^{-R\sin\theta} \cos\theta\, d\theta.$$

The new integral can be evaluated explicitly and evidently approaches 0 as $R \to \infty$.

Another approach is to give up the traditional use of circular-arc contours. For the Fresnel integrals (the more difficult case), take the contour to be the square with vertices at 0, $R$, $R + iR$, and $iR$ (initially with a small indentation at 0). The integrals along $(R, R + iR)$ and $(iR, R + iR)$ are

$$\int_0^R (R + iy)^{\lambda-1} e^{i(R+iy)}\, dy \text{ and } \int_0^R (x + iR)^{\lambda-1} e^{i(x+iR)}\, dx.$$

The sum of their absolute values does not exceed

$$\int_0^R (R^2 + y^2)^{(\lambda-1)/2} e^{-y}\, dy + \int_0^R (x^2 + R^2)^{(\lambda-1)/2} e^{-R}\, dx.$$

Since $R^2 + y^2 \geqslant R^2$ and $x^2 + R^2 \geqslant R^2$ (and $\lambda - 1 < 0$), the sum does not exceed

$$R^{\lambda-1}\int_0^R e^{-y}\, dy + e^{-R}R^{\lambda-1}\int_0^R dx = R^{\lambda-1}(1 - e^{-R}) + R^\lambda e^{-R} \to 0.$$

For integrals $e^{ix}R(x)\, dx$, the rectangular contour is used in [**1**] and [**3**].

Still another method [**2**] is to use the triangular contour with vertices $iR$, 0, and $R$. This gives a slightly more complicated equation for the part of the integral that approaches 0, but has only one integral instead of two to estimate.

**References**

1. L. V. Ahlfors, Complex Analysis, McGraw-Hill, New York, 1953, p. 127; 1969, p. 157.
2. R. P. Boas and E. Friedman, A simplification in certain contour integrals, this MONTHLY, 84 (1977) 467–468.
3. I. Stewart and D. Tall, Complex Analysis, Cambridge University Press, 1983, p. 223.

## PI

E. F. ASSMUS, JR.
*Department of Mathematics, Lehigh University, Bethlehem, PA 18015*

Most (and perhaps all) mathematicians involved in the teaching of calculus are asking themselves these days about the standard "techniques of integration" chapter: How much should be retained now that symbolic antidifferentiation is so readily available on home computers? Herbert Wilf described a more general question in his "The disk with the college education" [this MONTHLY, 89 (1982) 4–8], and I would like to suggest a general point of view by giving a particular answer to the "techniques of integration" question.

The point of view should, it seems to me, be that the home computer can free the student from

tedious drill and leave the teacher free to discuss the important conceptual aspects of integration. I take the consensus to be that change of variable, integration by parts, and the use of partial fractions must be taught: change of variable because of its theoretical importance and because it gives the student a chance to see how a hard problem becomes easy when looked at in the right way; integration by parts because of its constant use in both pure and applied mathematics as a theoretical tool; and partial fractions because one can show how only two transcendental functions (the logarithm and the arctangent) are necessary for integrating all rational functions.

Having agreed to this, what problems do we discuss to make our points? It seems to me they should be conceptual problems that the student can appreciate. I offer one example that I have found useful.

The students know how to compute the area, $\pi r^2$, and the perimeter, $2\pi r$, of a circle of radius $r$. The good students may even have wondered why this same mysterious $\pi$ appears in both formulae. For those who have not, this question gets their attention.

Assuming the students know how to use the definite integral to find area and arclength, one has

$$A = 4\int_0^r \sqrt{r^2 - x^2}\,dx, \quad P = 4\int_0^r \frac{r}{\sqrt{r^2 - x^2}}\,dx.$$

Now, the change of variable, $x = ru$, gives immediately that

$$A = \left\{4\int_0^1 \sqrt{1 - x^2}\,dx\right\} r^2$$

and

$$P = 2\left\{2\int_0^1 \frac{1}{\sqrt{1 - x^2}}\,dx\right\} r.$$

So, it is a question of relating

$$\int_0^1 \sqrt{1 - x^2}\,dx \text{ to } \int_0^1 \frac{1}{\sqrt{1 - x^2}}\,dx.$$

But the identity $1 = (1 - x^2) + x^2$ and integration by parts yield

$$\int_0^1 \frac{1}{\sqrt{1 - x^2}}\,dx = 2\int_0^1 \sqrt{1 - x^2}\,dx.$$

We've actually proved the existence of a number, $\pi$ say, with $A = \pi r^2$ and $P = 2\pi r$; we can calculate its approximate value by numerical integration; and we've done so using change of variable and integration by parts. Who could ask for anything more?

---

## 142. MISCELLANEA

Beauty is perfect, and perfection (such is human nature) holds our attention but for a little while. The mathematician who after seeing Phèdre asked: "*Qu'est-ce que ça prouve*?" was not such a fool as he has been generally made out.

—W. Somerset Maugham, *Cakes and Ale*.

---

### ANSWER TO PHOTOS ON PAGE 182

They are the brothers Alfred and Richard Brauer.

Let $g(x)$ be a polynomial with positive degree and integer coefficients which is irreducible over the rationals. Prove the following: (a) there exists an integer $n$ such that $g(n)$ is not a square; (b) for every positive integer $r$ there exists a positive integer $n$ such that $p\|g(n)$ for at least $r+1$ primes $p$. (Here $p\|g(n)$ means $p|g(n)$ but $p^2\nmid g(n)$.)

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Permutations and Derangements

E 2947 [1982, 334]. *Proposed by Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Canada.*

Let $n$ and $k$ be integers such that $0 \leqslant k \leqslant n$. Show that

$$\sum_{i=0}^{k}\binom{k}{i}D_{n-i}=\sum_{j=0}^{n-k}(-1)^j\binom{n-k}{j}(n-j)!$$

where $D_m = m!\,\Sigma_{r=0}^{m}(-1)^r/r!$ denotes the derangement number of $(1,2,\dots,m)$.

*Solution* I *by Michael Woltermann, Washington and Jefferson College.* Let $S=\{1,2,\dots,n\}$ and $T=\{1,2,\dots,n-k\}$. Then $\binom{k}{i}D_{n-i}$ is the number of permutations of $S$ which fix $i$ elements of $S\setminus T$ and derange the remaining elements. Thus $\Sigma_{i=0}^{k}\binom{k}{i}D_{n-i}$ is the number of permutations of $S$ which derange $T$. But by the principle of inclusion-exclusion, $\Sigma_{j=0}^{n-k}(-1)^j\binom{n-k}{j}(n-j)!$ is the number of permutations of $S$ fixing none of $1,\dots,n-k$, i.e., the number of permutations of $S$ which derange $T$.

*Solution* II *by C. S. Karuppan Chetty, Regional Engineering College, Tiruchirapalli, India.* Let $f(x)=(1-x)^{-1}e^{-x}$. The coefficient of $x^m$ in $f(x)$ is $D_m/m!$ and hence $D_m=f^{(m)}(0)$. By the Leibniz rule we have

$$\left[\frac{d^k}{dx^k}\{e^x f^{(n-k)}(x)\}\right]_{x=0}=\sum_{i=0}^{k}\binom{k}{i}D_{n-i}. \tag{1}$$

Since

$$f^{(n-k)}(x)=\sum_{j=0}^{n-k}(-1)^j\binom{n-k}{j}e^{-x}\left\{\frac{(n-k-j)!}{(1-x)^{n-k-j+1}}\right\},$$

the right-hand side of (1) is also equal to $\Sigma_{j=0}^{n-k}(-1)^j\binom{n-k}{j}(n-j)!$ and the result follows.

*Solution* III *by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* Integration by parts yields $D_{n-i}=\int_0^\infty(x-1)^{n-i}e^{-x}\,dx$, so

$$\sum_{i=0}^{k}\binom{k}{i}D_{n-i}=\int_0^\infty\sum_{i=0}^{k}\binom{k}{i}(x-1)^{n-i}e^{-x}\,dx=\int_0^\infty(x-1)^{n-k}x^ke^{-x}\,dx.$$

On the other hand, since $(n-j)!=\int_0^\infty x^{n-j}e^{-x}\,dx$,

$$\begin{aligned}\sum_{j=0}^{n-k}(-1)^j\binom{n-k}{j}(n-j)! &= \int_0^\infty x^k\sum_{j=0}^{n-k}(-1)^j\binom{n-k}{j}x^{n-k-j}e^{-x}\,dx\\ &= \int_0^\infty(x-1)^{n-k}x^ke^{-x}\,dx,\end{aligned}$$

as desired.

Also solved by M. Ašić (Yugoslavia), K. L. Bernstein, J. C. Binz (Switzerland), L. Carlitz, I. Gessel, S. H. Greene, V. Hernandez (Spain), J. G. Heuver, T. B. Kraus, C. Levesque and G. Lord (Canada), R. B. Nelson, I. Paasche (West Germany), U. Peled (Canada), O. G. Ruehr, J. Schwaiger (West Germany), K. Seyffarth (Canada), University of South Alabama Problem Group, J. Suck (West Germany), A. Varecza (Hungary), K. Williamson, P. Y. Wu (Republic of China), and the proposer.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by July 31, 1985. The solver's full post-office address should be on each sheet.*

6491. *Proposed by Jon Borwein, Dalhousie University, Canada.*

(a) Show that

$$4\pi = \frac{\sum_{n=-\infty}^{\infty} e^{-n^2\pi}}{\sum_{n=-\infty}^{\infty} n^2 e^{-n^2\pi}}.$$

(b) More generally, show that for each positive integer $k$

$$4\pi = \frac{\sum_{n=0}^{\infty} kr_k(n)\, e^{-n\pi}}{\sum_{n=0}^{\infty} nr_k(n)\, e^{-n\pi}},$$

where $r_k(n)$ is the number of distinct representations of $n$ as a sum of $k$ integral squares.

6492. *Proposed by Anatole Beck, University of Wisconsin-Madison.*

For which $\alpha$ is $f(x) = \int_0^x |t|^\alpha |\sin(1/t)|^{|1/t|}\, dt$ differentiable at $x = 0$?

## SOLUTIONS OF ADVANCED PROBLEMS

### A Probability Problem

6433 [1983, 402]. *Proposed by Edmund Butler, New Carrollton, MD.*

Let $X_1, X_2, \ldots, X_{n+1}$ be a sequence of independent random variables uniformly distributed on $[0,1]$. For any sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ of $\pm 1$'s let

$$P_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \Pr\{\operatorname{sign}(X_1 - X_2) = \varepsilon_1, \operatorname{sign}(X_2 - X_3) = \varepsilon_2, \ldots, \operatorname{sign}(X_n - X_{n+1}) = \varepsilon_n\}.$$

(a) Show that $M_n = \operatorname{Max}_{\varepsilon_j = \pm 1} P_n(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is assumed when the $\varepsilon$'s alternate in sign.
(b) Find $\lim_{n\to\infty} M_n^{1/n}$.

*Composite solution based on solution by L. E. Mattics, University of South Alabama, Mobile, Alabama to part* (a); *and solution by Alain Tissier, Montfermeil, France to part* (b). Let $E_n = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, $\varepsilon_i = \pm 1$, and let $A_{n+1}(E_n)$ be the set of rearrangements $y_1, y_2, \ldots, y_{n+1}$ of $1, 2, \ldots, n+1$ such that $\operatorname{sign}(y_{i+1} - y_i) = \varepsilon_i$, $i = 1, 2, \ldots, n$. Then

$$|A_{n+1}(E_n)| = |A_{n+1}(-E_n)| \quad \text{and} \quad P_n(E_n) = |A_{n+1}(E_n)|/(n+1)!.$$

Denote $|A_{n+1}(E_n)|$ by $C_{n+1}$ when $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ alternate in sign and set $C_0 = C_1 = 1$.

Assume it has been shown that $C_{m+1} \geqslant |A_{m+1}(E_m)|$ for all sequences $E_m$ with $1 \leqslant m \leqslant n$ (the initial case is trivial). To calculate the number of elements in $A_{n+2}(E_{n+1}) \cup A_{n+2}(-E_{n+1}) = K$, note that there are at most $\binom{n+1}{i-1}C_{i-1}C_{n+2-i}$ rearrangements $y_1, y_2, \ldots, y_{n+2}$ of $1, 2, \ldots, n+2$ in $K$ with $n+2$ in the $i$th position, and there are precisely that many if $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ alternate in sign. Hence

$$2|A_{n+2}(E_{n+1})| \leqslant \sum_{i=1}^{n+2} \binom{n+1}{i-1} C_{i-1}C_{n+2-i} = 2C_{n+2},$$

and this establishes (a) by induction.

To establish (b), note that $M_n = C_{n+1}/(n+1)!$ and set

$$f(x) = 1 + \sum_{n=0}^{\infty} M_n x^{n+1}.$$

Then the identity above yields $2f'(x) = f(x)^2 + 1$ so that

$$f(x) = \sec x + \tan x.$$

It is known that, for $|x| < \pi/2$,

$$\sec x = \sum_{n=0}^{\infty} \frac{|E_{2n}|}{(2n)!} x^{2n} \quad \text{and} \quad \tan x = \sum_{n=1}^{\infty} \frac{|B_{2n}|}{(2n)!}(4^{2n} - 4^n)x^{2n-1},$$

where the $E$'s are Euler numbers and the $B$'s are Bernoulli numbers. It is also known that

$$\frac{|E_{2n}|}{(2n)!} \sim 2\left(\frac{2}{\pi}\right)^{2n+1} \quad \text{and} \quad \frac{|B_{2n}|}{(2n)!} \sim \frac{2}{(2\pi)^{2n}}.$$

Since

$$M_{2n-2} = \frac{|B_{2n}|}{(2n)!}(4^{2n} - 4^n) \quad \text{and} \quad M_{2n-1} = \frac{|E_{2n}|}{(2n)!},$$

it follows that $M_n \sim 2(2/\pi)^{n+2}$ and hence that $\lim_{n\to\infty} M_n^{1/n} = 2/\pi$.

Also solved by Ira Gessel, Lajos Takács, and the proposer.

Ira Gessel pointed out that (a) is an immediate consequence of a result of I. Niven, A combinatorial problem of finite sequences, *Nieuw Arch. Wisk.* (3) 16 (1968), 116–123; and that D. André, Développements de séc $x$ et de tang $x$, *C. R. Acad. Sci. Paris*, 88 (1879), 965–967, showed that $f(x) = \tan x + \sec x$.

### A Periodic Sequence

6439 [1983, 569]. *Proposed by Morton Brown, University of Michigan.*

Let $\{a_n\}$ be a sequence of real numbers satisfying the relation $a_{n+1} = |a_n| - a_{n-1}$. Prove that $\{a_n\}$ is periodic with period 9.

*Solution by James F. Slifker, Florida International University, Miami, Florida.* The sequence contains some nonpositive terms, for if $a_n > 0$, $a_{n+1} > 0$ and $a_{n+2} > 0$, then $a_{n+3} = -a_n < 0$. In fact, the sequence contains two consecutive nonpositive terms, for if $a_n > 0$, $a_{n+1} \leqslant 0$ and $a_{n+2} > 0$, then $a_{n+5} = a_{n+1} + a_n < 0$ and $a_{n+6} = -a_n < 0$.

Suppose therefore that $a_m = -a \leqslant 0$ and $a_{m+1} = -b \leqslant 0$. Then the segment of the sequence starting at $a_m$ and ending at $a_{m+10}$ is

$$-a, -b, a+b, a+2b, b, -a-b, a, 2a+b, a+b, -a, -b.$$

Since $a_{m+9} = a_m$ and $a_{m+10} = a_{m+1}$, it follows that the sequence has period 9. Further, since $a_{m+4} = -a_{m+1}$ and $a_{m+5} = -a_{m+2}$, the period cannot be less than 9 except when every $a_n = 0$.

Also solved by the proposer and sixty-one others. The problem turned out to be rather elementary.

**A Rational Function Which Is a Polynomial**

6441 [1983, 569]. *Proposed by F. W. Schmidt and R. Simion, Southern Illinois University.*

Let $a_1 < a_2 < \cdots < a_n$ be $n$ distinct positive integers. Show that the rational function

$$\prod_{1 \le i < j \le n} \frac{x^{a_i} - x^{a_j}}{x^i - x^j}$$

is actually a polynomial.

*Solution by R. H. Jeurissen, Mathematics Institute, Nijmegen, The Netherlands.* It suffices to show that every zero of the denominator is also a zero of the numerator with at least equal multiplicity. Clearly 0 is a zero of the numerator with multiplicity $a_1(n-1) + a_2(n-2) + \cdots + 2a_{n-2} + a_{n-1}$, and this is at most its multiplicity in the denominator since $a_i \geqslant i$ for $i = 1, 2, \ldots, n-1$. Moreover all other zeros of the denominator are roots of unity. For $r = 1, 2, \ldots, n-1$, let $z_r$ be a primitive $r$th root of unity. Then $z_r$ is a zero of $x^{a_i} - x^{a_j}$ if and only if $a_j \equiv a_i \pmod r$. Suppose, for $s = 1, 2, \ldots, r$, that $k_s$ of the numbers $a_1, a_2, \ldots, a_n$ are in the residue class mod $r$ of $s$. Then the number of pairs $(i, j)$ with $1 \leqslant i < j \leqslant n$ for which $a_j \equiv a_i \pmod r$ equals

$$\sum_{s=1}^{r} \binom{k_s}{2} = \frac{1}{2} \sum_{s=1}^{r} k_s^2 - \frac{n}{2},$$

and so this is the multiplicity of $z_r$ as a zero of the numerator. The above expression, as a function of $k_1, k_2, \ldots, k_r$, has minimum value when $|k_s - k_t| \leqslant 1$ for $1 \leqslant s < t \leqslant r$, which is precisely what happens when $a_i = i$, since then

$$k_s = [n/r] + 1 \text{ for } 1 \leqslant s \leqslant n - r[n/r] \quad \text{and} \quad k_s = [n/r] \text{ otherwise.}$$

Also solved by C. Bandt & W. Tefera (Ethiopia), Mihàly Bencze (Romania), Aage Bondesen (Denmark), F. W. Dodd & L. E. Mattics, Ira Gessel, Robert Gilmer & Warren Nichols, Tadeusz Januszkiewicz & Ryszard Szwarc (Poland), O. P. Lossers (Netherlands), J. G. Mauldon, Pei Yuan Wu (Republic of China), D. Richman, Allen J. Schwenk, C. Wildhagen (Netherlands), and the proposers.

---

# REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Neyman*: *From Life*. By Constance Reid. Springer-Verlag, New York, 1982. 298 pp. $19.80.

SAMUEL KOTZ
*Department of Management Science & Statistics, University of Maryland, College Park, MD 20742*

Sir Robert Giffen begins one of the earliest textbooks on the subject, entitled simply *Statistics** with the following comment: "I propose to write a handbook on Statistics without giving a formal definition of the word. There are perhaps too many definitions in existence." He continues:

---

*As indicated on the front page, the volume was written about 1898–1900 and edited by Henry Higgs with the assistance of G. U. Yule and published by Macmillan, London, in 1913. G. U. Yule and M. G. Kendall's classical textbook, *Introduction to the Theory of Statistics*, originally authored by Yule in 1911 and then joined by Kendall in the 11th edition in 1937, may be familiar to many older readers of this MONTHLY.

"There has been much debate also on the question of whether Statistics is a distinct science or not.... It seems to be quite unnecessary to debate whether the whole field of Statistics thus dealt with or a portion of it can be treated as a distinct science... ." Giffen explains: "The main fact always is that... Statistics undoubtedly constitutes a branch of knowledge; that in the use of them, scientific method is indispensible; and that... the knowledge which is derived from Statistics in each study is frequently so separate that a distinct account can be given of each."

How relevant are these words, written some 85 years ago, to the debate on the role of statistics still evident in scientific circles? The number of definitions has increased quite substantially since Giffen's days. For a comprehensive collection of definitions of statistics, see the article by Walter F. Willcox (1935), a prominent American statistician. This list was extended by Nalimov in 1974, containing over 100 definitions in Russian, and was retranslated into an English edition in 1981. A comprehensive discussion of definitions of statistics is given in T. A. Bancroft's 1970 Presidential Address published in the Journal of the American Statistical Association in 1971.

Some 25 years after Giffen*, in response to a suggestion by E. R. Hedrick, Chairman of the American Mathematical Association's Dictionary Committee, to submit sample definitions for the Dictionary, a prominent mathematician, H. L. Rietz of the State University of Iowa,** presented a brief paper "to explain the meaning of Statistics and of the associated expressions: statistical data, statistical methods, theory of statistics, mathematical statistics, statistical probability and to suggest a list of terms and expressions from statistics, probability and insurance that should probably be included in the dictionary." Some quotes from his paper are very illuminating, both from the aspect of their substance as well as from an historical perspective. "Statistics is a comparatively new word. Its first occurrence in English seems to be in J. F. von Bielfeld's, *The Elements of Universal Erudition*, translated by W. Hooper, London 1770."*** "Statistics... gradually came to mean an exposition of the attributes of the state by numerical methods.... The word next came to denote the figures used in such descriptions.... This use of the word prevails at the present time, but the data may refer to the state or to any other subject... ." "The expression *statistical methods* means methods which are suitable for the description and characterization of statistical data.... A set of mathematical propositions that relate to statistical methods is often called *mathematical statistics*.... The general problem of mathematical statistics in its ideal form is to determine a system of drawings to be carried out with urns of fixed composition, in such a way that the results of the sets of drawings lead, with a high degree of probability, to a table of values identical with statistical data... ." "Mathematical statistics is thus one branch of the theory of probability... ." "The concept of statistical probability is involved whenever the properties of an aggregate are predicted or inferred by observations of a sample taken from the aggregate. Many such inferences are drawn by persons unfamiliar with mathematical statistics, and there is practically no doubt that many conclusions thus obtained are invalid." (Even though it was written over 60 years ago, how relevant this last sentence is to the current state of affairs in quantitative scientific practice.) "Mathematical statistics aims to establish criteria that give numerical values to the degrees of confidence to be placed in such inferences."

In *Current Problems of Mathematical Statistics*, a paper delivered at the International Congress of Mathematicians in 1954 in Amsterdam (over 30 years after Rietz's definitions), Jerzy Neyman† assesses the status of mathematical statistics in the 50's:

*This MONTHLY, vol. 29 (October 1922) 333–337.

**H. L. Rietz's seminal works on fundamental aspects of mathematical statistics appeared in, among other journals, Ann. Math. Statist. (1931–1939), Bull. Amer. Math. Soc. (1932) and Biometrika (1931–1932). (See Ann. Math. Statist., 1944, 15, p. 105, for a complete bibliography.) Rietz is also the author of Carus Mathematical Monograph No. 3, *Mathematical Statistics*.

***See, however, G. U. Yule (1905) for a different initial source.

†As pointed out in the book under review: "In 1954, Neyman and Tarski—Berkeley's "anti-Poles"—were two of five mathematicians from the U.S. who were invited to give featured addresses at the International Congress of Mathematicians."

> In general, the present stage of development of mathematical statistics may be compared with that of analysis in the epoch of Weierstrass. During the preceding decades, a very considerable number of special problems were solved without particular care for generality and rigor. The workers in the field were in too great a hurry to broaden the domain of research to bother with what they considered troublesome details. At the present moment, we observe the inevitable reaction. The basic concepts of the theory are being revised, the commonly accepted assertions are being examined in order to determine the exact domain of their validity and, just as in the times of Weierstrass and immediately thereafter, counterexamples are being constructed in order to show the insufficiency of the old proofs."

This perceptive assessment of mathematical statistics in 1954 should be a source of great satisfaction to a mathematical statistician in 1983. In less than 30 years, we have progressed from the era of K. Weirstrass (1815–1897), through the period of the last universalist, H. Poincaré (1854–1912), to the times of N. Bourbaki. By now, the rigor and generality of arguments in mathematical statistics have beeen well and routinely established; asymptotic results are proved under precise and verifiable "regularity conditions." "Troublesome" mathematical details do not, in general, present insurmountable difficulties to modern day mathematical statisticians, who can also resort to simulation and computer-oriented methodology. Whether the characterization of mathematical statistics as a branch of probability, valid in 1922, is fully justified today is questionable. Many branches of modern statistics, such as experimental design, regression analysis, sample surveys, classification and grouping techniques (in particular, clustering), multidimensional scaling and factor analysis, are at best weakly related to probability theory but are substantially mathematically oriented and can be appropriately considered as topics in *mathematical* statistics.

This substantial leap forward is primarily due to the efforts and accomplishments of Jerzy Neyman, a Russian-born Pole, teaching in London in the late thirties. On August 12, 1938, he got off the train at the Southern Pacific Station in Berkeley and for over 40 years, charted and substantially dominated the course of the rapid development of mathematical statistics, not only at Berkeley, but on the international scene as well. It should be noted that prior to his arrival at Berkeley, among other achievements, the indefatigable Jerzy Neyman originated the famous Neyman-Pearson Lemma which even today is considered the fundamental tool of hypothesis testing despite more recent appearances of competitive approaches. He put forward the theory of confidence intervals, one of the most important (albeit controversial) approaches to data analysis, contributed substantially to the theory of contagious distributions, which is very useful in biological and ecological applications, paved the way for a new dimension of modern sampling theory by developing sampling from stratified populations, and pioneered a new approach in the field of experimentation by devising a model for randomized experiments.

In her biography of Jerzy Neyman (1894–1981), who received the National Medal of Science for "laying the foundations of modern statistics and devising tests and procedures that have become essential parts of the knowledge of every statistician," Constance Reid, with relatively few technicalities and scientific minutiae, succeeds remarkably well in portraying Neyman's strong and colorful personality and his contributions to statistical methodology and applications. As W. H. Kruskal, past President of the American Statistical Association, writes in a letter to the author: "Your biography of Jerzy is a delight and a source of instruction; what a tour de force to present so clear a picture with so few technical excursions." The book is based on extensive interviews the author conducted with Neyman in the years 1978–1979, as well as personal interviews and written communications that the author diligently carried out with persons closely associated with Neyman in the U.S. and abroad. Notable among them was Egon Pearson whose joint groundbreaking work with Neyman in the thirties propelled and charted the course of Neyman's "complicated but not uninteresting" and highly productive life.

The skillful author presents a panorama of Neyman's life and activities for 85 years as she relates his birth on April 16, 1894, in Bendery, "the gate of Bessarabia," of Polish parents,

completion of his undergraduate studies in pure mathematics in 1917 in Kharkov in the Ukraine, his first job as statistician at the Polish National Institute in Bydgoszcz (Bromberg) from 1921–1923, his subsequent employment in Warsaw as an assistant at the University (1923), and as a special lecturer at the Central College of Agriculture. It was during his tenure in the College of Agriculture in 1925 that he received a fellowship to study statistics at University College in London with the "father of modern statistics," the famous Karl Pearson (1857–1936). We read with great interest the thrilling description of the development of his long friendship with Karl Pearson's son Egon, and the subsequent tension between them and their rival, the brilliant but apparently obnoxious Sir Ronald A. Fisher, his visits to Paris and his encounters with Borel and Lebesgue, his return to Poland and continued association and elaborate correspondence with Egon Pearson which resulted in the publication of their classical "joint statistical papers" on *Hypothesis Testing* during the years 1933–1938, and his return to England in early 1934 where he remained until 1938. In this year he accepted the offer from the University of California "to coordinate the work in Statistics in the University and especially to develop the subject on the mathematical side." The rest is history, presented by Ms. Reid in a most informative and entertaining manner, with great affection, enthusiasm and talent.

I would, however, hesitate to endorse this book as the sole, authoritative source for the history of modern statistics for the uninitiated. There is an obvious and sometimes unjustified overemphasis of Neyman's contributions (undoubtedly groundbreaking) at the expense of other American statisticians. As W. H. Kruskal points out in a private letter, "I see no mention of John Tukey anywhere, nor of George Box, although Joan Box's biography of Fisher has three index entries. The late Jimmie Savage appears twice in the Index but the text entries are brief and superficial." When asked in 1979 about his attitude toward the currently popular Bayesian view of statistical inference, Professor Neyman's reply to the author (page 274) was: "It does not interest me. I am interested in frequencies." This minor episode may serve as an indication of Neyman's "occasional refusal to pay attention to important scientific work done by other statisticians" (in the words of Professor Kruskal). The author's justifiable fascination with her "subject," especially his gallantry, charm, progressive and liberal thinking, devotion to colleagues and students causes her to underplay this aspect of his personality.

Unfortunately, a list of Neyman's contributions is not included in the book and the reviewer found somewhat annoying the author's habit of interrupting the smooth and absorbing chronological narration with light anecdotes and personally witnessed episodes in Neyman's life and routines in 1978 and 1979. Some of what Reid gives us in these excursions may seem suspiciously like padding.

Nevertheless, I recommend that anyone even remotely associated with statistics join the author in her interesting and enthusiastically sympathetic description of the life of a remarkable person and great scientist, whose dedication and unyielding determination elevated statistics to the status of a bona fide scientific and academic discipline and who was a major contributor to its foundation. To acquire a concurrent but somewhat different view of the course of the development of modern statistics, I would recommend J. Box-Fisher's biography [2] of her father, R. A. Fisher, to supplement Ms. Reid's volume.

A more accurate, broader and perhaps deeper history of statistics, with an emphasis on American contributions still awaits its author. However, we should all be grateful to Ms. Reid for stepping off into unchartered territory and laying the foundation for such a work.

*Personal Note*. Since 1978, the reviewer, with Professor N. L. Johnson of the University of North Carolina, has been engaged in editing an 8-volume Encyclopedia of Statistical Sciences. In 1980, when the editors approached the late Professor Neyman to contribute to the ESS, his response was most encouraging; regrettably, the transitory nature of human life prevented him from fulfilling his commitment. We were gratified, however, to receive from him several unsolicited letters advising us on the work and suggesting entries with the names of suitable authors appropriate for our endeavor. We are proud of this association with Neyman during his last years.

**References**

1. T. A. Bancroft, The American Statistical Association: A single scientific and educational community, J. Amer. Statist. Assoc., 66 (1971) 7–12.
2. J. Box-Fisher, R. A. Fisher—The Life of a Scientist, Wiley, New York, 1978.
3. R. Giffen, Statistics, Macmillan, London (posthumous edition), 1913.
4. W. H. Kruskal, Letter to Constance Reid, Department of Statistics, University of Chicago, 1983.
5. V. V. Nalimov, Probability Model of Languages, Nauka, Moscow, Appendix I, 1974, pp. 242–263. Collection of definitions of Statistics. (Contains over 100 definitions from 1749 to 1971.) English translation: In the Labyrinths of Language: A Mathematician's Journey (edited by R. G. Colodny), ISI Press, Philadelphia, 1981, Appendix I, pp. 207–226.
6. J. Neyman, Current problems of mathematical statistics, Proceed. Intern. Congress of Mathematicians, 1954, Amsterdam, Sept. 2–Sept. 9, vol. 1, North-Holland, Amsterdam, 1957, pp. 349–370.
7. H. L. Rietz, Statistics in Mathematical Encyclopedic Dictionary, this MONTHLY, 29 (1922), 333–337.
8. W. F. Willcox, Definitions of Statistics, ISI Rev. 3, no. 4 (1935) 388–399.
9. G. U. Yule, The introduction of the words "statistics," "statistical" into the English language, J. Roy. Statist. Soc., 68 (1905) 391–396.


---

*Differential Equations: Classical to Controlled.* By Dahlard L. Lukes. Mathematics in Science and Engineering, vol. 162. Academic Press, New York, 1982. xiii + 322 pp. $37.50.


EMILIO ROXIN
*Department of Mathematics, University of Rhode Island, Kingston, RI 02881*


The subject of differential equations always played a central role in the development of the infinitesimal calculus. Indeed, an obvious generalization leads from the problem of "antiderivative" $\dot{x} = dx/dt = f(t)$ to the differential equation $(E)\dot{x} = dx/dt = f(t, x)$. From the very beginning, Isaac Newton made differential equations an almost universal tool to model the physical world, and the success was so impressive that it opened the doors for a great development and understanding of the laws of Nature. Differential equations have been applied not only to the physical world (say, physics and astronomy), but also to much more "human" sciences such as economics, physiology, population dynamics, epidemiology and management sciences. They have even had an impact on philosophical questions like causality and determinism.

It would be interesting to make a thorough study of the emphasis and style of the most representative books on differential equations over the years. This would show, besides the technical advances, the changing attitudes of mathematicians and the different expectations of what this theory was supposed to accomplish. This review certainly cannot present such an exhaustive study, but will contribute just a few remarks of this nature and address, in particular, the introduction of control theory into the subject of differential equations.

In the beginning, the naive approach to the problem posed by the differential equation $(E)$ was "to find a function $x = x(t)$ which satisfies $(E)$ identically in $t$". From this starting point many useful methods have been developed, including classification of the equations into different types, and also numerous tricks, to be used skillfully, in finding the unknown function.

The next step was to discover that the correct question to ask was to solve the "initial value problem," and to give conditions assuring existence and uniqueness of the solution. These questions found their proper place within the theory in the early 19th century work of Cauchy. They were significantly improved within the present century by the introduction of the "Carathéodory conditions," which allow $f(t, x)$ in $(E)$ to be just measurable (hence discontinuous) in $t$. This is of utmost importance in the theory of optimal control, where "bang-bang" controls are in some sense "best."

Linear equations of the forms $\dot{x} = Ax$ and $\dot{x} = Ax + b$, and in particular equations and systems of equations with constant coefficients, are an ideal ground where with relatively modest

effort one can obtain many explicit results. Physically speaking, linearity means the principle of superposition, which is basic for a simple formulation of the laws of elasticity, vibrations, electromagnetic circuits and so on. Its successful use was the basis on which the growing technology of the early 20th century flourished. It was the time of "linear engineering," when nonlinearities, if too strong to be neglected, were taken into account by tables, diagrams, or, as a last resort, by a conveniently generous safety factor. That it is possible to achieve a lot with such methods is shown by the power plants, electrical networks and machinery, bridges, skyscrapers and so on, which were built within this period.

Meanwhile, the "mathematical" interest expanded into the analytic theory. Ordinary differential equations were considered within the framework of (real or complex) analytic functions. This gave us power series expansions and the study of singular points. Periodic systems of o.d.e.'s inherit some of the nice properties of the systems with constant coefficients. Floquet theory is already classical, but the interest in periodic systems and periodic solutions carried over into the area of optimal control theory, where it is a problem of current research interest.

Solutions of o.d.e.'s were characterized as "dynamical systems," a name which may carry some application-oriented flavor, and soon a bridge was built into the most highly abstract regions of topology ("topological dynamics").

The impossibility of "solving" most differential equations in closed form made it imperative to look for other methods to attack the problem. Numerical methods were developed early and, in the present era of the computer, have received a tremendous boost. Limitation of space is the reason for not addressing this subject here and for leaving out several other important areas such as boundary value problems, eigenfunctions, etc.

For mildly nonlinear equations, a representation of the form $\dot{x}$ = linear part + $\varepsilon$ times perturbation, and a corresponding development of the solution in powers of $\varepsilon$ proved very useful. We find these in the work of H. Poincaré. These power series also gave birth to the very curious "asymptotic series," which can be very good for numerical computations even in the case when they are divergent! Related to this approach is also the "method of averaging," where a nonlinear oscillating system can be treated as a linear one, using some "average" parameters.

The methods above still try to obtain approximations to the solution of a differential system. The "second" or "direct" method of Lyapunov, in contrast, uses inequalities in order to obtain upper bounds needed to ascertain stability properties of the solutions. This method was introduced at the turn of the century and revived in the thirties. This led to the consideration of all kinds of differential inequalities.

The post-World War II era saw a growth in interest in nonlinear differential equations in several directions. In particular stability theory and nonlinear phenomena were in the spotlight.

Topological properties of the solutions, an idea going back to I. Bendixson, became suddenly of interest. This subject, which includes limit sets, invariant sets, and index theory, became known as "geometric theory of differential equations" in the work of Lefschetz which had a great impact. It included the birth of modern optimal control theory.

The genesis of control theory can be traced back in history. James Watt's centrifugal regulator is usually cited, but earlier regulating systems can be found in Antiquity and in the Middle Ages. In World War II this became the subject of "servomechanisms." Controls acted upon a system via "feedback." The systems considered were mostly linear, but a "linear theory" of a sustained oscillation (as is the case in any radio oscillator circuit) is a self-contradiction: if the system were really linear, the oscillations would necessarily grow to infinity or die out to zero. Hence nonlinearities are essential. Extremely nonlinear feedback, implemented by on-off relays (="bang-bang" controls), proved to be the cheapest, and in some mathematical sense, the optimal control for many systems. Bushaw proved this for the harmonic oscillators, $\dot{x} = y$, $\dot{y} = -x + u(t)$, with control $u$ bounded $-1 \leqslant u \leqslant 1$ but otherwise arbitrary; the state $(x, y)$ is supposed to be brought from a given initial state to the rest position $(0,0)$ in minimum time. The optimal solution turns out to be a control of the "bang-bang" type: it only takes values $\pm 1$, and the whole problem reduces to finding the switching points.

The subject of optimal control mushroomed in the fifties and became well established and rounded off in the sixties. The famous "Pontryagin's principle" was established. Its roots, of course, were in the classical calculus of variations. Parallel developments led R. Isaacs to create the "differential games," which are control systems with two controllers with opposed objective, and R. Bellman to develop "dynamic programming," which is the discrete counterpart of optimal control.

The present book of D. L. Lukes brings a new and innovative look to several areas of the theory of ordinary differential equations. Matrix manipulations for solving linear systems are developed systematically (the "ABC" method). Time varying linear systems of a particular kind ("commutative" systems) are shown to share many nice properties of the systems with constant coefficients. Finally, control systems get a respectable place. In particular, questions like controllability and stabilizability by linear feedback, which are of practical importance, are well treated. Many of the given results are due to the author or at least show his personal flavor.

While this book does not claim to be the last word in research accomplishments on the subject, represented by other more encyclopedic works, it certainly is in line with an up-to-date outlook. It is a very refreshing addition to the existing literature.

---

*From the Calculus to Set Theory, 1630–1910: An Introductory History*. Edited by I. Grattan-Guinness. Duckworth, London, England, 1980. 306 pp.

ROBIN E. RIDER
*History of Science and Technology Program, The Bancroft Library, University of California, Berkeley, CA 94720*

For many years the standard guide to the history of the calculus was Carl Boyer's 1939 work, republished in 1949 as *The History of the Calculus and Its Conceptual Development*, and reissued in a popular paperback edition by Dover in 1959. Since that date, the contours of scholarship in the history of mathematics have begun to change. An encyclopedic guide to the previous literature, edited by the late Kenneth O. May and published in 1973, identified areas of strength and weakness in the historiography and offered guidance to those undertaking research in the field. New journals for the history of mathematics have been founded; the most accessible to English-speaking readers is *Historia mathematica*, also a product of May's energy and enthusiasm for the subject. In addition to the traditional genres of encyclopedic histories and collections of biographical anecdotes, readers can now choose among intellectual biographies rich in both personal and mathematical detail[1], analyses of the institutional milieux and social contexts in which mathematicians operated[2], detailed studies of topics in mathematics[3], and new editions of the works and papers of individual mathematicians[4]. The authors of these recent studies have taken advantage of both published and unpublished sources, and have in some cases expanded their purview to include minor figures as well as major heroes in the history of mathematics.

Six historians of mathematics, all of whom have participated in this resurgence of interest and activity, have joined forces in *From the Calculus to Set Theory, 1630–1910*. The book covers more than the standard territory of the history of calculus. It is intended as an historical introduction to a set of interrelated developments in mathematics: the origins of the differential and integral calculus in the 17th and 18th centuries; the elaboration of mathematical analysis in the 19th and early 20th centuries; and the reexamination of the foundations of mathematics, especially the progress of set theory and mathematical logic in the latter part of the 19th century. The level of exposition, sometimes heavily mathematical, presupposes considerable mathematical knowledge. As the editor cautions, "this is *not* the place from which the mathematics can be *learnt*" (p. 7). The book is thus directed at readers with some training in mathematical analysis and set theory, especially upper-division undergraduates and graduate students.

The six chapters are each sufficiently self-contained to stand alone; any can profitably be read by those with interests in a particular mathematical domain. Chapter 1, by Kirsti Møller Pedersen, guides the reader through the thicket of techniques and special methods for solving problems of rectification and quadrature, tangents, and maxima and minima of curves during the period 1630–1660. Chapter 2, by H. J. M. Bos, explores the work of Newton, Leibniz, and Leibniz' 18th-century followers in the creation and application of a coherent, unified differential and integral calculus. Bos concludes with an admirable summary of the influential work of Euler, the starting point for the next chapter. In Chapter 3, Ivor Grattan-Guinness traces the route by which mathematical analysis developed from Euler through Cauchy and the Weierstrass school to about 1880. Chapter 4, by Thomas Hawkins, goes over the same chronological ground, but with special attention to the development of the integral and its theory by such mathematicians as Cauchy, Riemann, and Lebesgue. In Chapter 5, Joseph Dauben summarizes his research on Georg Cantor and the genesis of set theory. General questions of the foundations of mathematics from 1870 to 1910 are discussed by Robert Bunn in the final chapter.

To survey mathematical research over the course of nearly three centuries demands selectivity, and the contributors to this book have in general chosen to focus on the well-known names in the history of mathematics. What is too often left out is information and interpretation regarding the historical context in which the great mathematicians operated. For example, both Bos and Dauben treat briefly related philosophical questions and the importance of personality. Readers interested in pursuing these questions further, however, will have to turn elsewhere—to such works as A. Rupert Hall's book on the prolonged and vitriolic dispute engendered by competing claims to the invention of the calculus, and Dauben's book-length biography of Cantor, which interweaves psycho-emotional, mathematical, and philosophical themes[5]. Likewise, Bos and Grattan-Guinness both mention in passing the links between mathematical analysis and other sciences, especially the rational mechanics of the 18th century and the mathematical physics of the 19th. Physical problems stimulated investigations in mathematical analysis and helped shape the form of results; many of the key figures in the history of mathematical analysis played important roles in the development of physics as well. Grattan-Guinness' book (with J. Ravetz) on Fourier and the study of heat explores the connection between mathematics and physics in the work of one individual at a time when French mathematical methods profoundly influenced physical science[6]. A growing number of studies on mathematical physics in the 19th century consider the relation between Continental mathematical methods and British physics[7]. The book under review also gives little indication of the variety of opinions and the vigor of ongoing debate regarding the origins of a rigorous formulation of calculus. For example, Grattan-Guinness' assessment of Cauchy's work differs markedly from that of Judith Grabiner. To learn about this debate, however, one must turn to their books on the foundations of mathematical analysis[8].

Such shortcomings are more striking in light of the book's announced purpose. The editor and contributors, who are conscious of the tendency in mathematics to stress "the accumulation of mathematical knowledge," wish instead to explore "the *growth* of mathematical *understanding*, the appreciation of why a mathematical theory developed and took its *form*, and not merely that it does have its *content*" (p. 3). To downplay the historiographical issues, the social context, the philosophical and personal dimensions, the connections with other sciences, is to undermine the purpose of the book. Inclusion of developments in set theory and the foundations of mathematics is a commendable revision to the usual history of the calculus. But an attempt to balance this largely internalist account with considerations of the broader historical context would be more welcome. Mathematicians as well as other scientists (and historians of science) should know, and would like to know, more about the evolution of so important a concept as the calculus.

**Notes**

1. E.g., Michael S. Mahoney, The Mathematical Career of Pierre de Fermat 1601–1665, Princeton, 1973.
2. Paul L. Rose, The Italian Renaissance of Mathematics: Studies on Humanists and Mathematicians from Petrarch to Galileo, Geneva, 1975.

3. Thomas Hawkins' articles on group representation theory in Arch. History Exact Sci., 7 (1971) 142–170; 8 (1972) 243–287.

4. The volumes of correspondence in the Euler Opera Omnia (Basel), and Newton, The Mathematical Papers, D. T. Whiteside (Editor), 8 vols., Cambridge, 1967–1981.

5. A. Rupert Hall, Philosophers at War: The Quarrel Between Newton and Leibniz, New York, 1980; Joseph Dauben, Georg Cantor, Cambridge, Mass., 1979.

6. Ivor Grattan-Guinness and J. R. Ravetz, Joseph Fourier 1768–1830, Cambridge, Mass., 1972.

7. See the essay review by M. Norton Wise, The Maxwell literature and British dynamical theory, Historical Studies in the Physical Sciences, 13 (1982) 175–205.

8. I. Grattan-Guinness, The Development of the Foundations of Mathematical Analysis from Euler to Riemann, Cambridge, Mass., 1970; Judith V. Grabiner, The Origins of Cauchy's Rigorous Calculus, Cambridge, Mass., 1981.

---

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

Since the appearance of my note [**Le**] on Weierstrass' Theorem, I have been informed of several sources of proofs that are very similar to mine. These proofs appear in [**C**], [**F**], [**Lah**], and [**Lam**]. In addition, I have found a short discussion of how the proof can be obtained in [**G**].

R. Streit points out that a generalization of the Bernstein proof appears in [**H**], Theorem 10.4.3 and (19.2.14).

I find it interesting that this proof of the most important theorem in approximation theory seems to be better known to probabilists than approximation theorists.

### References

[**C**] Y. Chow and H. Teicher, Probability Theory, Springer-Verlag, New York, 1978.

[**F**] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, Wiley, New York, 1971.

[**G**] R. R. Goldberg, Methods of Real Analysis, Wiley, New York, 1976.

[**H**] E. Hille and R. S. Phillips, Functional Analysis and Semi-groups, AMS Colloquium Publications, vol. XXXI, 1957.

[**Lah**] R. G. Laha and V. K. Rohatgi, Probability Theory, Wiley, New York, 1979.

[**Lam**] J. Lamperti, Probability: A Survey of the Mathematical Theory, W. A. Benjamin, New York, 1966.

[**Le**] K. M. Levasseur, A Probabilistic Proof of the Weierstrass Approximation Theorem, this MONTHLY, 91 (1984) 249–250.

Kenneth M. Levasseur
Department of Mathematics
University of Lowell
Lowell, MA 01854

---

Editor:

Recently, in a discussion of estate tax law in a publication by Commerce Clearing House (a well-known and highly respected organization, whose writings on the tax laws are second in authority only to the Federal Code itself), I encountered the following remarkable passage:

Suppose, it is hypothesized, that a decedent has specified in his will that after certain specific bequests are made from his estate, the balance is to be used to pay taxes and the residue given to charity.

The author notes that, since charitable bequests are deductible from the taxable estate, the amount of the charitable contribution will affect the tax; but that the amount of the tax will in turn determine the size of the charitable contribution. Neither can be determined until the other is known. This, he concludes, is a mathematical problem of great difficulty.

Great difficulty indeed! Problems of this type appear routinely in the Rhind papyrus no less, and techniques for solving them have been known for over 38 centuries! Even the most elementary algebra is not really required—such problems can be solved by a clever method known as "false positioning". (For details, consult J. R. Newman, "The Rhind Papyrus" in *The World of Mathematics*, Vol. I, pp. 170–178.)

The CCH author goes on to suggest Trial and Error as an acceptable method of solution and even cites an Internal Revenue Service form on which the requisite arithmetic can be reported!

To the average person, even the average educated person, a problem such as the one above involving two unknown quantities that are mutually dependent on each other presents a challenge of overwhelming magnitude. The sad part is that many of these people have probably studied elementary algebra and know how to solve the problem—they just don't *know* they know.

F. W. Luttmann
Department of Mathematics
Sonoma State University
Rohnert Park, CA 94928

---

## 143. MISCELLANEA

### Some More Clerihews

Evariste Galois
Was far from bourgeois;
His ideas political and mathematical
Verged on the radical.

Georg Cantor
Would often wander
In the general vicinity
Of infinity.

Kurt Gödel
Cleared many a hurdle
To show that mathematicians
Cannot have omniscience.

Nicolas Bourbaki
Is quite a mystery to me;
It is difficult to write a verse on
A pluralistic person.

—Samuel Bates
Department of Mathematics
Northwestern University
Evanston, IL 60201

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 4 April 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, Michigan State University, East Lansing, MI 48824. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is \$22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are \$22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is \$55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# BULGARIAN SOLITAIRE

ETHAN AKIN* AND MORTON DAVIS
*Department of Mathematics, CUNY, The City College, New York, NY 10031*

Blast Martin Gardner! There you are, minding your own business, and Scientific American arrives like a virus. All else forgotten, you must struggle with infection by one of his fascinating problems. In the August 1983 issue he introduced us to Bulgarian solitaire:

> Form a pile of 45 cards, then divide it into as many piles as you like, with an arbitrary number of cards in each pile. You may leave it as a single pile of 45 or cut it into two, three or more piles, cutting anywhere you want, including 44 cuts to make 45 piles of one card each. Now keep repeating the following procedure. Take one card from each pile and place all the removed cards on the table to make a new pile. The piles need not be in a row. Just put them anywhere. Repeat the procedure to form another pile, and keep doing it.
>
> As the structure of the piles keeps changing in irregular ways it seems unlikely you will reach a state where there will be just one pile with one card, one pile with two cards, one with three and so on to one with nine cards. If you should reach this improbable state, without getting trapped in loops that keep returning the game to a previous state, the game must end, because now the state cannot change. Repeating the procedure leaves the cards in exactly the same consecutive state as before. It turns out, surprisingly, that regardless of the initial state of the game, you are sure to reach the consecutive state in a finite number of moves.

He then quotes an amazing theorem of Jørgen Brandt that Bulgarian solitaire always terminates in the consecutive sequence $\{1,2,\ldots,k\}$ providing, of course, that the number of cards in the deck is the triangular number $t_k \equiv 1 + 2 + \cdots + k$. But what if the number of cards is not triangular? As we later discovered, Brandt answered this question as well. Since a deck has only finitely many layouts, the play of Bulgarian solitaire must eventually cycle (we include equilibrium as a cycle of period 1). In his elegant, but terse, paper [**1**] Brandt characterizes and counts all cycles for any given deck size. The result quoted by Gardner—*Brandt's Equilibrium Theorem*, we will call it—follows because the consecutive layout is the only cycle when the deck size is a triangular number.

In elaborating Brandt's results we hope to share with the reader our delight in Bulgarian solitaire viewed as a simple sort of dynamical system.

In observing a game of Bulgarian solitaire we need a way of describing the state of the deck at any time. The simplest way to describe a layout of the deck is to list the sizes of the various stacks. Because we disregard the arrangement of the stacks we can make the listing decreasing (or more precisely, nonincreasing). In this way we get the *rank order* description:

$$S_k \equiv \text{the size of the } k\text{th stack listed in decreasing order of size.}$$

---

*Ethan Akin*: After my undergraduate work at City College of New York, I did my Ph.D at Princeton in piecewise-linear topology. Since returning to City College I have worked primarily on applications to biology and economics. I am also interested in philosophy, history and horses. But I never, never touch problems in combinatorics. (Well, hardly ever.)

*Morton Davis*: I received my Ph.D. at the University of California (Berkeley) in 1961. My dissertation, most of my research papers, and an introductory book were all addressed to my primary interest: game theory. I have also worked in artificial intelligence (computer learning) and on financial models. I am now finishing my 19th year at the City College of New York.

*Support by NIH Grant 1RO1 GM 32153-01 GEN.

We get the *frequency distribution* by counting, for each whole number $j$, how many stacks contain exactly $j$ cards:

$$N_j \equiv \text{the number of stacks of size } j.$$

Finally, the *residual distribution* counts, for each whole number $i$, how many stacks contain at least $i$ cards:

$$Q_i \equiv \text{the number of stacks of size} \geqslant i.$$

The latter two modes of description are related by the equations:

(1) $$Q_i = \sum_{j \geqslant i} N_j, \qquad N_j = Q_j - Q_{j+1}.$$

The residual and rank order descriptions are related by a useful sort of geometrical duality which Gardner mentions. Notice first that $Q$ is an infinite sequence of natural numbers. It is nonincreasing, with finitely many nonzero terms followed by an infinite string of zeroes. We can state this somewhat formally by: $Q \in \mathscr{L}$, where our "state space," $\mathscr{L}$, is the set defined by:

$$\mathscr{L} \equiv \{ P = (P_1, P_2, \dots) : P_1 \geqslant P_2 \geqslant \cdots, \quad P_i \in \mathbb{Z}^+ \text{ for all } i \text{ and } P_k = 0 \text{ for some } k \}.$$

Here $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$ is the set of natural numbers.

We can also regard the rank order sequence $S$ as a member of $\mathscr{L}$ by attaching a terminal string of zeroes. In any example we will describe the $S$ or $Q$ sequence by listing the initial nonzero part and omit the zeroes.

For example, with $S = 8, 3, 3, 2, 1, 1$ we can picture the layout of the deck as in Fig. 1 (upper left). Instead of stacks of cards we see stacks of dots with each dot marking a point whose $x$ and $y$ coordinates are whole numbers. So a point with whole number coordinates $(k, i)$ is marked with a dot precisely when $S_k \geqslant i$.

In general, for any sequence $P$ in $\mathscr{L}$ we define the *shadow* of $P$ by

(2) $$\mathrm{Sh}(P) = \{(i, j) : i, j \in W \text{ and } P_i \geqslant j\},$$

where $W = \{1, 2, 3, \dots\}$ is the set of whole numbers. Clearly, we can recover $P$ by observing its shadow. For $P_i$ we count the shadowed points with $x$-coordinate $i$, or alternately look at the $y$-coordinate of the highest point with $x$-coordinate $i$:

(3) $$P_i = \max\{ j : (i, j) \in \mathrm{Sh}(P)\}.$$

By convention, we define the maximum over the empty set to be zero ($\max \varnothing = 0$).

Now suppose $S$ is the rank order sequence for some layout, as in Fig. 1. To say that $(k, i)$ is in the shadow of $S$ means that $S_k \geqslant i$. Because the sequence is nonincreasing, this means that all of the first $k$ stacks have size at least $i$, i.e., $Q_i \geqslant k$. This argument goes the other way as well and shows that $S_k \geqslant i$ (i.e. $(k, i) \in \mathrm{Sh}(S)$) if and only if $Q_i \geqslant k$ (i.e. $(i, k) \in \mathrm{Sh}(Q)$). In other words, that shadow of the residual distribution $Q$ is obtained from the shadow of the rank order sequence $S$ by just transposing the $x$ and $y$ coordinates. Thus, in Fig. 1 (upper right) we see that shadow of the residual distribution $Q = 6, 4, 3, 1, 1, 1, 1, 1$.

For any sequence $P$ in $\mathscr{L}$ we can define the *dual* sequence $P^*$ in $\mathscr{L}$ exactly this way. For $P_k^*$ we count the $P$-shadowed points with $y$ coordinate $k$ or look at the $x$-coordinate of the rightmost point with $y$-coordinate $k$:

(4) $$P_k^* = \max\{ i : P_i \geqslant k \},$$

where, again, $\max \varnothing \equiv 0$. It is clear that in general $\mathrm{Sh}(P^*)$ is obtained from $\mathrm{Sh}(P)$ by transposing coordinates. This shows that $P^{**} = P$.

We can summarize all this by saying that the rank order sequence $S$ and the residual sequence $Q$ are dual to one another, where duality in $\mathscr{L}$ is defined via (4). This duality is well known to

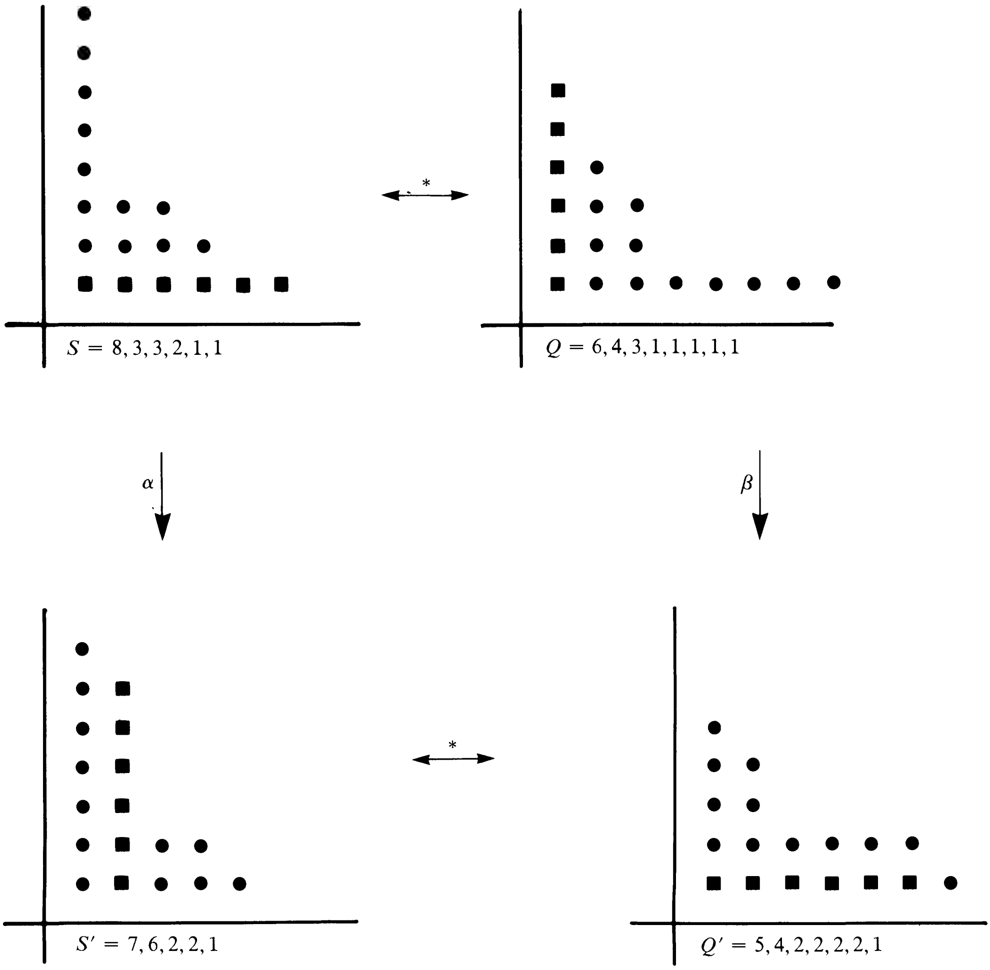



FIG. 1

those who count things other than cards. For example, in [3] May exploits this duality in discussing abundance of biological species on islands.

Now let's play some Bulgarian solitaire. We denote by $\alpha$ the map which associates to the rank order sequence of a layout the rank order of the new layout after a round of play. Fig. 1 (left) illustrates what happens. Pull one card from each stack—we use the American riverboat convention of extracting it from the bottom of each stack—and fit the resulting new stack in an appropriate place among the remains of the old stacks. We get $S' = \alpha(S)$. When $S = 8,3,3,2,1,1$, $S' = 7,6,2,2,1$ as shown. The size of the new stack is the number of old stacks, in this case: 6. In general, for any layout:

$$\max\{k\colon S_k > 0\} = \sum_j N_j = Q_1 = \text{the number of stacks.} \tag{5}$$

We denote by $\beta$ the dual of $\alpha$ which associates to the residual sequence of a layout the residual sequence of the new layout after one round. The first column is pulled off, rotated 90° to get a row of dots and then this new row is pushed underneath the other columns after they are moved left one place. This describes $Q' = \beta(Q)$. In Fig. 1 (right) with $Q = 6,4,3,1,1,1,1,1$, $Q' =$

5, 4, 2, 2, 2, 2, 1 which is dual to $S' = 7, 6, 2, 2, 1$.

By sufficient contemplation of Fig. 1 the reader may be able to convince himself that this messy description is indeed dual to $\alpha$. But the question tends to arise at this point—if it hasn't earlier—why bother with the dual at all? The residual distribution is the least intuitive way of describing a layout so why introduce it in the first place?

The answer is that while—as the descriptions suggest—the physical play of the game is best understood using $S$, it turns out that the analysis is best done using $Q$. This is because the formula for the map $\alpha$ is somewhat messy. But the formula for the map $\beta$ is easy to work with. Let's derive it by looking first at the frequency distribution $N$.

After one round of play the number of new stacks of size $j$ is the number of old stacks of size $j+1$, except that there is one additional new stack of size $Q_1$ (= the old number of stacks by (5)). This is because we have taken one card from each old stack and put them together to form one new stack. Notice that the old stacks of size 1 are gone.

We can write this as:

$$N_j' = N_{j+1} + \delta(Q_1)_j. \tag{6}$$

The second term in (6) is the Kronecker delta defined by

$$\delta(i)_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

If we now define

$$\chi(\leqslant k)_i \equiv \sum_{j \geqslant i} \delta(k)_j = \begin{cases} 1, & i \leqslant k, \\ 0, & i > k, \end{cases}$$

then we can get the formula for $Q_i' = \beta(Q)_i$ by summing (6) on the $j$'s $\geqslant i$ (see (1)):

$$\beta(Q)_i = Q_{i+1} + \chi(\leqslant Q_1)_i. \tag{7}$$

So, for example, the new stack number is $\beta(Q)_1 = Q_2 + 1$: one more than the number of old stacks of size at least 2. Notice that formula (7) corresponds to the description we gave before: the columns of $\mathrm{Sh}(Q)$ are shoved left one place after the first column is removed and then the first $Q_1$ columns are jacked up by 1 as the $Q_1$ row is shoved underneath.

Instead of a formula for $\alpha$, we can just use duality to say:

$$\alpha(S) = \beta(S^*)^*, \tag{8}$$

i.e., for analytic purposes, we go from the rank order description to the dual residual sequence, apply $\beta$ to get the new residual sequence and take the dual to get the new rank order.

We emphasize again that we will keep going back and forth: $\alpha$ for intuition and $\beta$ for analysis. ("Heaven for climate, Hell for company"—M. Twain.)

As we play, the total number of cards in the deck doesn't change. This number is given by the equations:

$$\sum_k S_k = \sum_j jN_j = \sum_i Q_i = \text{the size of the deck.} \tag{9}$$

The first two formulas are easy to check. The third can be derived from the second using (1) but it is easier to use duality and note that for either $S$ or $Q$ the sum is just the size of the shadow. In general, for any $P \in \mathscr{L}$ we define the *size* of $P$:

$$\|P\| = \sum_i P_i = \text{the number of points in } \mathrm{Sh}(P). \tag{10}$$

As an exercise, we prove that $\beta$ preserves size by using the analytic definition, (7):

$$\|\beta(Q)\| = \sum_{i=1}^{\infty} Q_{i+1} + \sum_{i=1}^{\infty} \chi(\leqslant Q_1)_i.$$

The first sum is $\|Q\| - Q_1$ while the second consists of $Q_1$ ones and so sums to $Q_1$. Hence,

$$\|\beta(Q)\| = \|Q\|.$$

Our original set of sequences $\mathscr{L}$ can be cut up into different levels by defining

$$\mathscr{L}_r = \{P \in \mathscr{L}: \|P\| = r\}.$$

The finite set $\mathscr{L}_r$ is exactly the set of all rank order sequences (or alternatively the set of all residual frequency sequences) of layouts for a deck of size $r$.

This raises another objection. In discussing Brandt's paper we mentioned that it was the finiteness of the set of layouts of a deck which made the cycling inevitable. In a roundabout way, we introduced the infinite set $\mathscr{L}$ and then sliced it back up into finite sets. Why not just fix the deck size $r$ and restrict attention to $\mathscr{L}_r$? The answer to this is the moral of our mathematical story.

The representation of some operation by a map from a set to itself, e.g., $\alpha$: $\mathscr{L} \to \mathscr{L}$ or $\beta$: $\mathscr{L} \to \mathscr{L}$, is the prototype of a dynamical system model. Iterating the mapping corresponds to applying the operation repeatedly, starting from an initial point. The first step in studying such a dynamical system is to look for structures on the set, i.e., ways of measuring, combining or comparing elements of the set, which are preserved by the map in some fashion. Why $\mathscr{L}$? Because $\mathscr{L}$ is rich in structures which are only partly apparent in each separate $\mathscr{L}_r$. This explains why our analysis—once we get to it—all proceeds by comparing sequences at different levels of $\mathscr{L}$.

To illustrate this idea of comparing different levels we examine the fate, under $\beta$, of an extra 1 included on top of some column of $Q$:

(11)

$$\beta(Q + \delta(1)) = \beta(Q) + \delta(Q_1 + 1), \qquad \beta(Q + \delta(k)) = \beta(Q) + \delta(k - 1) \qquad (\text{if } k > 1).$$

The proof of (11) is an exercise for the reader in the use of (7) and the Kronecker delta notation. To see what it says go back to Fig. 1 (right) and put your finger on top of some column of $Q$. If you start over some column beyond the first ($k > 0$), then after the operation your finger has moved along with its column one step to the left (to the $k - 1$ place). If your finger started out on the first column, then it now ends up underneath the new $Q_1 + 1$ (in this case $= 7$) column, lifting it up one place. A technical point for the wary: adding on an extra 1 may take the sequence out of $\mathscr{L}$ because $Q + \delta(k)$ might not be monotonic. However, $\beta$ can be defined via (7) for arbitrary sequences of natural numbers and its consequent (11) is true in general.

In addition to the duality relation and the measurement of size, the structures on $\mathscr{L}$ that we use are a partial ordering and a closely related metric.

For $P, Q$ in $\mathscr{L}$ we define a *partial ordering* (i.e., a reflexive, antisymmetric and transitive relation) by:

(12) $$P \geqslant Q \Leftrightarrow P_i \geqslant Q_i \qquad i = 1, 2, \ldots.$$

So $P \geqslant Q$ means for each $i$, $P_i$ is at least as high as $Q_i$. From this it is clear that

(13) $$P \geqslant Q \Leftrightarrow \mathrm{Sh}(P) \text{ contains } \mathrm{Sh}(Q) \text{ as a subset.}$$

We define the *distance* between $P$ and $Q$ by

(14) $$\|P - Q\| = \sum_{i=1}^{\infty} |P_i - Q_i|.$$

We will see below that the metric, too, can be described using shadows.

Let us collect together the relations among the four pieces of structure (size, duality, order and metric) that we have defined on $\mathscr{L}$.

PROPOSITION 1. *Let $P, Q \in \mathscr{L}$.*

(a) $\|P^*\| = \|P\|$, $\|P - Q\| = \|P^* - Q^*\|$, *and $P \geqslant Q$ if and only if $P^* \geqslant Q^*$.*

(b) $\|P - Q\| \geqslant \|P\| - \|Q\|$ *with equality if and only if $P \geqslant Q$.*

(c) $\|P - Q\| = $ *minimum* $\{\|\tilde{P}\| - \|\tilde{Q}\|\}$, *where $\tilde{P}$ and $\tilde{Q}$ vary over all members of $\mathscr{L}$ such that $\tilde{P} \geqslant P, Q$ and $\tilde{Q} \leqslant P, Q$.*

*Proof*. Recall that the shadows of $P$ and $P^*$ are just transposes of one another. $\|P\|$ and $\|P^*\|$ are the sizes of the respective shadows and so they are equal. Similarly, $\mathrm{Sh}(P)$ contains $\mathrm{Sh}(Q)$ if and only if $\mathrm{Sh}(P^*)$ contains $\mathrm{Sh}(Q^*)$. The distance result also follows by looking at shadows as we shall see.

For (b), notice that $|P_i - Q_i| \geqslant P_i - Q_i$ with equality if and only if $P_i \geqslant Q_i$. Now sum on $i$ and notice that equality between the sums requires equality at every $i$.

For (c) we build two new elements of $\mathscr{L}$: $P \vee Q$ and $P \wedge Q$ by defining for each $i$:

$$(P \vee Q)_i = \max(P_i, Q_i),$$
$$(P \wedge Q)_i = \min(P_i, Q_i). \tag{15}$$

Clearly, $P \vee Q \geqslant P, Q$ and $P \wedge Q \leqslant P, Q$. Because $|P_i - Q_i| = \max(P_i, Q_i) - \min(P_i, Q_i)$, we can sum on $i$ to get:

$$\|P - Q\| = \|P \vee Q\| - \|P \wedge Q\|. \tag{16}$$

It is easy to check that

$$\mathrm{Sh}(P \vee Q) = \mathrm{Sh}(P) \cup \mathrm{Sh}(Q),$$
$$\mathrm{Sh}(P \wedge Q) = \mathrm{Sh}(P) \cap \mathrm{Sh}(Q).$$

From these equations and (16) we get the shadow interpretation of distance:

$$\|P - Q\| = \text{the number of points in } \mathrm{Sh}(P) \text{ or } \mathrm{Sh}(Q) \text{ but not in both.} \tag{17}$$

Because this count is preserved when we transpose coordinates, it follows that $\|P^* - Q^*\| = \|P - Q\|$, completing (a)

(c) also follows because $\tilde{P} \geqslant P, Q$ implies $\tilde{P} \geqslant P \vee Q$, and so $\|\tilde{P}\| \geqslant \|P \vee Q\|$. Similarly, $\tilde{Q} \leqslant P, Q$ implies $\|\tilde{Q}\| \leqslant \|P \wedge Q\|$. Hence, $\|\tilde{P}\| - \|\tilde{Q}\| \geqslant \|P - Q\|$ by (16). QED

REMARK. This result shows the relationship between the ordering and the metric. Given the concept of size we can get the ordering from the metric by (b) or go the other way by (c).

Because duality preserves order, we can say that a layout of some deck is greater than or equal to a layout (of a possible different deck) if the corresponding rank order sequences are related by $\geqslant$ or, equivalently, if the corresponding residual frequency sequences are related by $\geqslant$. This leads to a nice picture of the ordering which we formalize as:

PROPOSITION 2. *Start with a layout of cards. Now deal out additional cards onto existing stacks or deal out some new stacks or both. The new layout is greater than the old layout. Conversely, we can get to any greater layout by dealing out new cards in this fashion.*

*Proof*. The new deal does not decrease the number of stacks of size $\geqslant i$. So $P_i \geqslant Q_i$ for all $i$, where $P$ and $Q$ are the residual sequences for the two layouts.

Conversely, if one layout is greater than another, then $\mathrm{Sh}(T)$ contains $\mathrm{Sh}(S)$, where $T$ and $S$ are the rank order sequences. So we can build $\mathrm{Sh}(T)$ from $\mathrm{Sh}(S)$ by adjoining new points, i.e., by dealing out cards. QED

In particular, notice that the restriction of the ordering to a single level is trivial, i.e., two layouts of the same deck can be compared by the ordering only when they are equal. This is what we meant when we said that some of the rich structure of $\mathscr{L}$ is lost when we restrict attention to $\mathscr{L}_r$.

Now we are ready to relate the structure on $\mathscr{L}$ to our dynamical system:

THEOREM 3. Let $P, Q \in \mathscr{L}$.

(a) $\|P\| = \|\beta(P)\| = \|\alpha(P)\|$.

(b) $\|\alpha(P) - \alpha(Q)\| = \|\beta(P) - \beta(Q)\| \leqslant \|P - Q\|$.

(c) *If* $P \geqslant Q$, *then* $\alpha(P) \geqslant \alpha(Q)$ *and* $\beta(P) \geqslant \beta(Q)$.

*Proof.* The results for $\alpha$ follow from the results for $\beta$ by equation (8) because duality preserves size, distance and ordering.

We have already proved (a) and the formal proof of (b) is very similar. From (7) it is clear that

$$\|\beta(P) - \beta(Q)\| \leqslant \sum_{i=1}^{\infty} |P_{i+1} - Q_{i+1}| + \sum_{i=1}^{\infty} |\chi(\leqslant P_1)_i - \chi(\leqslant Q_1)_i|.$$

The first sum is $\|P - Q\| - |P_1 - Q_1|$.

Now suppose that $P_1 \geqslant Q_1$. Then the terms of the second sum are zero except when $Q_1 < i \leqslant P_1$, because outside this range $\chi(\leqslant Q_1)_i$ and $\chi(\leqslant P_1)_i$ are both 0 or both 1. Inside the range

$$|\chi(\leqslant P_1)_i - \chi(\leqslant Q_1)_i| = \chi(\leqslant P_1)_i = 1.$$

Thus, the second sum is $P_1 - Q_1 = |P_1 - Q_1|$ and (b) follows.

For (c) notice that $P \geqslant Q$ implies $P_{i+1} \geqslant Q_{i+1}$ and $\chi(\leqslant P_1)_i \geqslant \chi(\leqslant Q_1)_i$ for all $i$. So from (7): $\beta(P)_i \geqslant \beta(Q)_i$ for all $i$.

As (c) is our most important result we give a conceptual proof using Proposition 2.

Suppose that the layout for $Q$ consists of stacks of blue cards. By the Proposition you can get the layout for $P$ by extracting some red cards from your sleeve and dealing them out onto the existing layout. Now in playing a round of solitaire make sure that from any stack which contains blue cards the one you pick to remove is blue (e.g., pull the card from the bottom of each stack). In the new stack put all the blue cards on the bottom. The new combined layout is the same as though you played a round with old blue and then dealt out some red cards on top. So by Proposition 2 again the new layouts are related by $\geqslant$. QED

REMARK. While we proved (b) and (c) independently they are actually equivalent in the presence of (a), i.e., if $\beta\colon \mathscr{L} \to \mathscr{L}$ is any map preserving size, then $\beta$ preserves order if and only if it does not increase distance. For example, suppose (a) and (b) are true and $P \geqslant Q$. By Proposition 1(b) and assumption (b) here we have:

$$\|P\| - \|Q\| = \|P - Q\| \geqslant \|\beta(P) - \beta(Q)\| \geqslant \|\beta(P)\| - \|\beta(Q)\| = \|P\| - \|Q\|.$$

Hence, the inequalities are equalities and $\beta(P) \geqslant \beta(Q)$ by Proposition 1(b) again. Similarly, (a) and (c) imply (b) via Proposition 1(c).

At last, we are ready to release the energy stored up by all this work. To the cycles!

Recall that if $r = t_k = k + k - 1 + \cdots + 1 = \frac{1}{2}k(k+1)$ and $P$ lies in $\mathscr{L}_r$, then Brandt's Equilibrium Theorem says that $\alpha^n(P)$ eventually equals the consecutive sequence $B(k)$, where

$$B(k) = k, k-1, k-2, \ldots, 2, 1, 0, 0, \ldots. \tag{18}$$

Because $\alpha(B(k)) = B(k)$, this implies that the operation terminates at $B(k)$ after a finite number of iterates.

Notice that $B(k)$ is self-dual, i.e., $B(k)^* = B(k)$, because the shadow is symmetric (see Fig. 2):

$$\mathrm{Sh}(B(k)) = \{(i, j)\colon i + j \leqslant k + 1\}.$$

For the equilibrium in $\mathscr{L}_r$, $B(k)$ is at once the rank order sequence and the residual sequence. So $\beta^n(P)$ is eventually equal to $B(k)$ as well, and $\beta(B(k)) = B(k)$.

Now suppose that $P$ lies in $\mathscr{L}_r$ but $r$ is not triangular: $t_{k-1} < r < t_k$. We use what might be called the method of "two big guys hustling a little guy out the door".

We can increase $P$ to get $\tilde{Q}$ in $\mathscr{L}$ with $\tilde{Q} \geqslant P$ and $\|\tilde{Q}\| = t_k$ (deal out $t_k - r$ cards). Also we can decrease $P$ to get $Q$ in $\mathscr{L}$ with $P \geqslant Q$ and $\|Q\| = t_{k-1}$ (lift up $r - t_{k-1}$ cards).

By the Equilibrium Theorem, there exists a large enough $n$ that $\alpha^n(\tilde{Q})$ and $\beta^n(\tilde{Q})$ are $B(k)$, and that $\alpha^n(Q)$ and $\beta^n(Q)$ are $B(k-1)$. Now because $\alpha$ and $\beta$ preserve order (Theorem 3(c)):

$$\alpha^n(P) \leqslant \alpha^n(\tilde{Q}) = B(k), \qquad \alpha^n(P) \geqslant \alpha^n(Q) = B(k-1),$$

and similarly for $\beta^n(P)$. So we have:

$$B(k-1) \leqslant \alpha^n(P), \beta^n(P) \leqslant B(k). \tag{19}$$

These inequalities exactly describe the cyclic elements of $\mathscr{L}_r$ as we now show.

THEOREM 4. *With $B(k)$, $B(k-1)$ the consecutive sequences defined by* (18), *let* $\mathscr{C}^{(k)} = \{P \in \mathscr{L}: B(k-1) \leqslant P \leqslant B(k)\}$.

(a) *The set $\mathscr{C}^{(k)}$ is closed under duality and the maps $\alpha$ and $\beta$, i.e., $P$ in $\mathscr{C}^{(k)}$ implies $P^*$, $\alpha(P)$ and $\beta(P)$ lie in $\mathscr{C}^{(k)}$.*

(b) *Every element of $\mathscr{C}^{(k)}$ is cyclic. In fact, $P \in \mathscr{C}^{(k)}$ implies*

$$P = \alpha^k(P) = \beta^k(P). \tag{20}$$

*Proof.* (a) is true because $B(k) = B(k)^* = \alpha(B(k)) = \beta(B(k))$ and because $*$, $\alpha$ and $\beta$ preserve $\geqslant$. For example,

$$B(k) = \beta(B(k)) \geqslant \beta(P) \geqslant \beta(B(k-1)) = B(k-1)$$

proves: $P \in \mathscr{C}^{(k)} \Rightarrow \beta(P) \in \mathscr{C}^{(k)}$.

Now we prove (b) and in the process describe $\mathscr{C}^{(k)}$ in some detail.

Notice first that $B(k)_i - B(k-1)_i$ is 1 for $i \leqslant k$ and is 0 afterwards, i.e.,

$$B(k) - B(k-1) = \chi(\leqslant k).$$

Hence, if $P$ lies between $B(k-1)$ and $B(k)$, then the difference $P_i - B(k-1)_i$ is either 0 or 1 for $i$ up to $k$ and is 0 afterwards. So if we define

$$\Delta_i = P_i - B(k-1)_i, \qquad i = 1, 2, \ldots, k,$$

then $\Delta$ is a function from $\{1, \ldots, k\}$ to the two-point set $\{0, 1\}$.

Conversely, if $\Delta$: $\{1, \ldots, k\} \to \{0, 1\}$ is any such function and we extend the definition by $\Delta_i = 0$ for $i > k$, then $P \in \mathscr{C}^{(k)}$, where

$$P = B(k-1) + \Delta. \tag{21}$$

This exhibits a one-to-one, onto correspondence between the set $\mathscr{C}^{(k)}$ and this set of functions which we denote $\{0, 1\}^k$.

Now we define $\tilde{\beta}$: $\{0, 1\}^k \to \{0, 1\}^k$ by

$$\tilde{\beta}(\Delta)_i = \begin{cases} \Delta_{i+1}, & i < k, \\ \Delta_1, & i = k. \end{cases} \tag{22}$$

Notice that 1 is congruent to $k + 1$ modulo $k$. So $\tilde{\beta}(\Delta)_i$ is obtained by evaluating $\Delta$ at $i + 1$ reduced back to $\{1, \ldots, k\}$ modulo $k$. Iterating $n$ times, $\tilde{\beta}^n(\Delta)_i$ is $\Delta$ evaluated at $i + n$ reduced modulo $k$. In particular, $i + k$ congruent to $i \bmod k$ implies:

$$\tilde{\beta}^k(\Delta) = \Delta. \tag{23}$$

The proof is completed by showing that, under the correspondence between $\mathscr{C}^{(k)}$ and $\{0, 1\}^k$, $\beta$ corresponds to $\tilde{\beta}$, i.e.,

$$\beta(B(k-1) + \Delta) = B(k-1) + \tilde{\beta}(\Delta). \tag{24}$$

To prove this with $P = B(k-1) + \Delta$, we want to check that

$$\beta(P)_i = B(k-1)_{i+1} + \Delta_{i+1} + \chi(\leqslant P_1)_i$$

is the same as

$$B(k-1)_i + \tilde{\beta}(\Delta)_i.$$

For $i < k$, $P_1 \geqslant B(k-1)_1 = k - 1$ implies

$$B(k-1)_{i+1} + \chi(\leqslant P_1)_i = B(k-1)_{i+1} + 1 = B(k-1)_i.$$

For $i = k$,

$$B(k-1)_{k+1} = \Delta_{k+1} = 0 = B(k-1)_k,$$

while $\chi(\leqslant P_1)_k$ is 0 if $P_1 = k - 1$ and is 1 if $P_1 = k$, i.e., $\chi(\leqslant P_1)_k = \Delta_1$. Finally, if $i > k$ everything is 0. QED

Fig. 2 illustrates all this. Think of $\Delta$ as a list of 0's and 1's placed on the $k$ whole number points of the line $x + y = k + 1$. The 1's mark the points of $\mathrm{Sh}(P) - \mathrm{Sh}(B(k-1))$. If $r = \|P\|$, then there are $r - t_{k-1}$ such points. In particular, $P = B(k-1)$ corresponds to the list of all 0's and $P = B(k)$ corresponds to the list of all 1's.

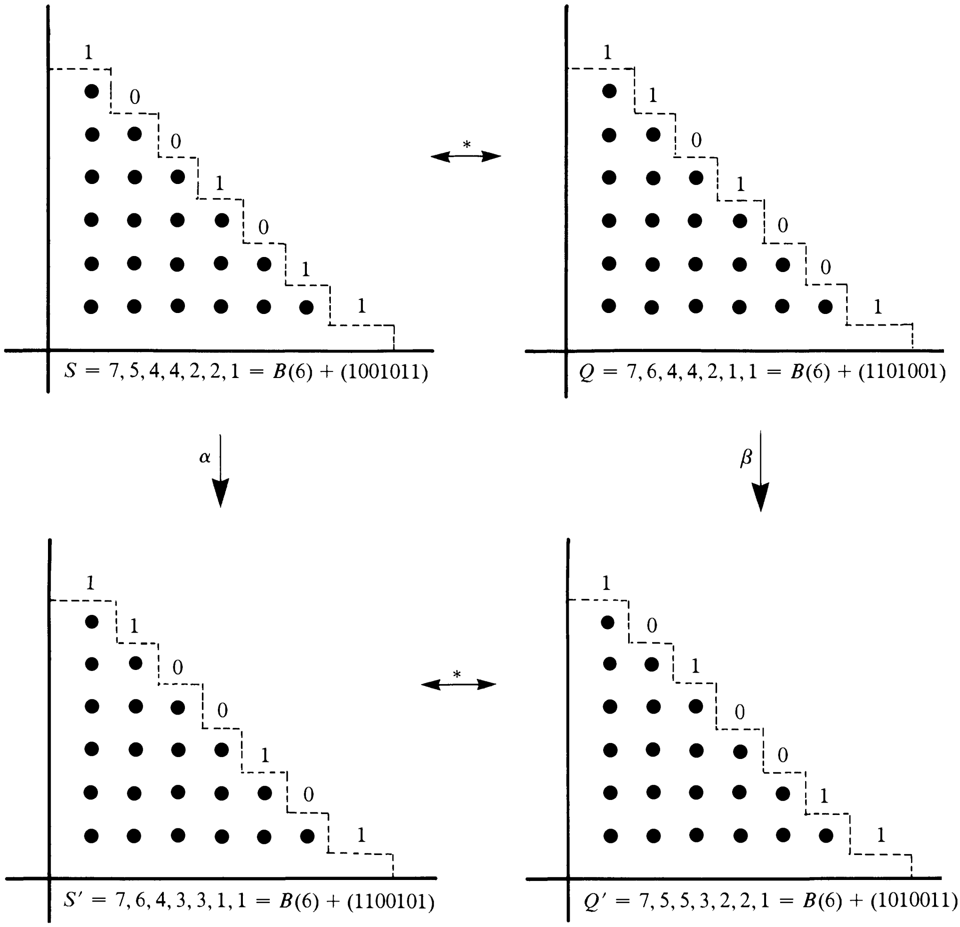



FIG. 2

The behavior of $\beta$ is illustrated by Fig. 2 (right). Think of $\mathrm{Sh}(B(k-1))$ upon which the 0's and 1's are standing as a staircase, but one of the paradoxical sort that occur in the prints of M. C. Escher. As $\beta$ is applied everyone moves up one step. The 0 or 1 which was originally on top ($i = 1$ column) takes a step and appears at the bottom ($i = k$ column). Applying this cyclic permutation $k$ times we get the identity, i.e., after $k$ steps everyone (and everyzero) is back in his original place.

For completeness we have illustrated $\alpha$ applied to the dual in Fig 2 (left). Notice that the dual

consists of the same list of 0's and 1's but in reverse order. $\alpha$ on $\mathscr{C}^{(k)}$ is the inverse of $\beta$. It corresponds to a step down the Escher staircase.

Theorem 4 identifies certain elements of $\mathscr{L}$ as cyclic. We proved (19) by assuming Brandt's Equilibrium Theorem. Thus, if $r = \|Q\|$ is between $t_{k-1}$ and $t_k$, then $\beta^n(Q)$ eventually lies in $\mathscr{C}^{(k)}$. In particular, the elements identified by Theorem 4 are all of the cyclic elements. We now reprove this result without assuming the Equilibrium Theorem and obtain that Theorem as a Corollary.

THEOREM 5. *Define* $\mathscr{C} = \mathscr{C}^{(2)} \cup \mathscr{C}^{(3)} \cup \cdots$, *i.e.*,

$$\mathscr{C} = \{ P \in \mathscr{L} \colon B(k-1) \leqslant P \leqslant B(k) \text{ for some } k = 2,3,\dots \}.$$

*If* $P \in \mathscr{L}$, *then* $\beta^n(P)$ *and* $\alpha^n(P)$ *lie in* $\mathscr{C}$ *for sufficiently large* $n$. *In particular,* $\mathscr{C}$ *consists of all the cyclic elements of* $\mathscr{L}$.

*Proof*. We prove that $\beta^n(P)$ eventually lies in $\mathscr{C}$ by induction on $\|P\|$. Duality implies the result for $\alpha$, and the rest of Theorem 5 then follows from Theorem 4.

Earlier we were able to bound $P$ from above and below by elements whose fates were known. Now we only have information from below. One of the big guys is no longer there so the other gets a bit closer and uses some judo.

If $\|P\| = 1$, then the result is trivial because $B(1)$ is the only member of $\mathscr{L}_1$.

Now if $\|P\| = r$ with $t_{k-1} < r \leqslant t_k$, we assume the result for $r-1$ ($\geqslant t_{k-1}$) and prove it for $r$. Choose $Q$ with $P \geqslant Q$ and $\|Q\| = r-1$ (i.e., lift one card from the corresponding layout). By inductive hypothesis there exists $n$ such that $\beta^n(Q) \in \mathscr{C}^{(k)}$ and so

$$\beta^n(Q) = B(k-1) + \Delta,$$

where $\Delta \in \{0,1\}^k$ as in the proof of Theorem 4.

Because $\beta$ preserves the ordering $\beta^n(P) \geqslant \beta^n(Q)$. Because $\beta$ preserves size $\|\beta^n(P)\| = r$ and $\|\beta^n(Q)\| = r-1$. So if we let $R = \beta^n(P)$, then $R$ differs from $\beta^n(Q)$ by the addition of 1 in some place:

$$R = B(k-1) + \Delta + \delta(j)$$

for some $j$.

Think of $\beta^n(Q) = B(k-1) + \Delta$ as a line of 0's and 1's standing on the Escher staircase. $R$ is built by the addition of a new 1 which hops aboard at position $j$. To keep track of this intruding 1 let us name it: Ralph.

What happens now depends on the character of the place $j$ at which Ralph has arrived. There are three kinds of places $i$:

$$i \text{ is empty} \qquad (i \leqslant k \text{ and } \Delta_i = 0):$$

If $j$ is an empty spot, then $\Delta + \delta(j)$ is a member of $\{0,1\}^k$ and so $R$ lies in $\mathscr{C}^{(k)}$.

$$i \text{ is occupied} \qquad (i \leqslant k \text{ and } \Delta_i = 1):$$

If Ralph arrives at an occupied spot, he hops on top of the 1 who is already there. $\Delta + \delta(j)$ takes on the value 2 at $j$ and so is not in $\{0,1\}^k$. $R$ is not in $\mathscr{C}^{(k)}$.

$$i \text{ is the waiting room} \qquad (i = k+1):$$

If Ralph arrives at $j = k+1$, then again $\Delta + \delta(j)$ is not in $\{0,1\}^k$ because it takes on a nonzero value (namely 1) at $j = k+1$. Again $R$ is not in $\mathscr{C}^{(k)}$.

Notice that if $j$ were greater than $k+1$, then $R_j = 1$ and $R_{k+1} = 0$ would contradict monotonicity of $R$ ($R$ is in $\mathscr{L}$). So these are the only possibilities.

So if Ralph arrived at an empty spot, then $R$ is in $\mathscr{C}^{(k)}$ and all is well. Otherwise, Ralph is sitting on top of another 1 on the staircase or is not on the staircase at all but on the $k+1$ position, just beyond the stairs, which we have labelled the *waiting room*. Because $R$ is not yet in

$\mathscr{C}^{(k)}$ we must continue to apply $\beta$. The proof is completed by using the following:

CLAIM. (waiting room) $\beta^k(B(k-1)+\Delta+\delta(k+1)) = B(k-1)+\Delta+\delta(1)$.
(occupied spot) If $\Delta_j = 1$, then $\beta^k(B(k-1)+\Delta+\delta(j)) = B(k-1)+\Delta+\delta(j+1)$.

Suppose that $R$ is not in $\mathscr{C}^{(k)}$, i.e., Ralph's position $j$ is not an empty one. The claim says that after $k$ iterations of $\beta$ the original underlying list of 0's and 1's, $\Delta$, is restored, but Ralph has moved. If he was originally at an occupied position $j$, then he is now one step down at $j+1$, including the possibility $(j=k)$ that he has moved out to the waiting room. If he was originally in the waiting room $(j=k+1)$, then he is now at the top of the stairs (at position 1).

Now if this new position is empty, then $\beta^k(R)$ is in $\mathscr{C}^{(k)}$. If not, we apply $\beta^k$ again. Ralph keeps moving rightwards (or clockwise if we include the leap from $k+1$ back to 1) one step at a time, sampling each position in turn until he finds an empty spot. Then the process terminates with entry into $\mathscr{C}^{(k)}$. Notice that there are certainly empty positions somewhere because $\|\beta^n(Q)\| = r-1 < t_k$ and so $\Delta$ does not consist entirely of 1's.

To prove the Claim we begin with a single application of $\beta$. By the exercise (11) (which we sneaked in for just this purpose):

$$\beta(R) = \beta(B(k-1)+\Delta+\delta(j)) = B(k-1)+\tilde{\beta}(\Delta)+\delta(z), \tag{25}$$
$$\text{where } z = j-1 \text{ if } j>1 \quad\text{and}\quad z = k+\Delta_1 \text{ if } j=1.$$

Here we have used (24) to evaluate $\beta(B(k-1)+\Delta)$ and we have used $(\beta(k-1)+\Delta)_1 + 1 = k + \Delta_1$ for the $j=1$ case.

The first half of the Claim $(j=k+1)$ follows immediately by iterating (25) $k$ times and using $\tilde{\beta}^k(\Delta) = \Delta$.

Now suppose $j \leqslant k$ and $\Delta_j = 1$. After $j-1$ iterates of (25), note that $z = j-(j-1) = 1$:

$$\beta^{j-1}(R) = B(k-1)+\tilde{\beta}^{j-1}(\Delta)+\delta(1)$$

and $\tilde{\beta}^{j-1}(\Delta)_1 = \Delta_j = 1$. So applying (25) again, we note that $z = k+\tilde{\beta}^{j-1}(\Delta)_1 = k+1$ and so:

$$\beta^{j}(R) = B(k-1)+\tilde{\beta}^{j}(\Delta)+\delta(k+1).$$

Now apply (25) $k-j$ times again. This time $z = k+1-(k-j) = j+1$:

$$\beta^{k}(R) = B(k-1)+\tilde{\beta}^{k}(\Delta)+\delta(j+1).$$

Finally, $\tilde{\beta}^k(\Delta) = \Delta$ completes the proof of the Claim and so, of the Theorem.

To see what is happening suppose that the 1 originally in position $j$ is named Quincy. As Quincy climbs the stairs during successive applications of $\beta$, Ralph rides along on his shoulders. At last, Quincy arrives at the top (position 1) with Ralph aboard. Then Quincy takes the magic Escher step which brings him back to the bottom of the stairs (position $k$) and in the process chucks Ralph out into the waiting room (position $k+1$). With the next application of $\beta$, Quincy moves up one step (to position $k-1$) and Ralph hops into the position just behind Quincy. If this place in line is empty, we are in $\mathscr{C}^{(k)}$. If not, then Ralph is now aboard the 1 behind Quincy and goes for another ride. QED

From the proof it is easy to construct sequences which take a long time to stabilize. Let

$$Q = k+1, k-1, k-2, \ldots, 3, 2, 0, 0.$$

$Q$ is $B(k+1)$ with $k$ and 1 deleted and so

$$\|Q\| = t_{k+1} - k - 1 = t_k.$$

$Q = B(k-1)+\Delta+\delta(1)$, where $\Delta = (1,1,\ldots,1,0)$. Thus, $\beta^k$ will have to be iterated $k-1$ times before the incoming 1 reaches the unoccupied position at the end, i.e., $Q$ requires $k(k-1)$ iterations of $\beta$ to reach $B(k)$. Gardner mentions the conjecture that $k(k-1)$ is the maximum number of iterations needed to reach $B(k)$ from any initial $P$ in $\mathscr{L}_{t_k}$. The recursive estimates

yielded by the proof above are much cruder because of the iterates needed to reach $R$ first. However, it is reported that the conjecture has been proved by Igusa.

The rank order sequence for the layout with residual sequence $Q$ is given by:

$$S = k-1, k-1, k-2, k-3, \dots, 3, 2, 1, 1.$$

So $S$ requires $k(k-1)$ iterations of $\alpha$ to reach $B(k)$.

The Equilibrium Theorem follows from Theorem 5 by observing that if $r = t_k$, then $\mathscr{L}_r \cap \mathscr{C} = \{B(k)\}$. The only cycling element at level $t_k$ is the equilibrium $B(k)$.

If $t_{k-1} \leqslant r \leqslant t_k$, then we can write $r = t_{k-1} + u$ with $0 \leqslant u \leqslant t_k - t_{k-1} = k$. By using the $\Delta$ representation we see that the cardinality of $\mathscr{L}_r \cap \mathscr{C}$, i.e., the number of cycling elements of size $r$, is given by the binomial coefficient $\binom{k}{u}$ counting the number of ways of choosing a subset of size $u$ from a set of size $k$. The subset is the 1-positions on the staircase.

More interesting is the question of the number of disjoint cycles. For example, if $r = t_{k-1} + 1$, $\mathscr{L}_r \cap \mathscr{C}$ consists of $k$ elements corresponding to $\Delta = (1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, \dots, 0, 1)$. But these all lie on a single cycle of period $k$. Similarly, if $r = t_k - 1$, $\mathscr{L}_r \cap \mathscr{C}$ consists of a single cycle of period $k$. However, if $t_{k-1} + 1 < r < t_k - 1$, i.e., $1 < u < k - 1$ (which requires $k \geqslant 4$), then there is always more than one cycle. For example, if $k = 4$ and $u = 2$, then $\Delta = (1, 0, 1, 0)$ and $\tilde{\Delta} = (1, 1, 0, 0)$ correspond to two elements of $\mathscr{L}_8 \cap \mathscr{C}$ generating the two different cycles:

$$Q = 4, 2, 2 \qquad (S = 3, 3, 1, 1)$$

and

$$\tilde{Q} = 4, 3, 1 \qquad (\tilde{S} = 3, 2, 2, 1).$$

We now give a proof of Brandt's formula for the number of disjoint cycles. Some familiarity with elementary number theory is needed for this sidetrip.

THEOREM 6. *Let $r = t_{k-1} + u$ with $0 < u \leqslant k$ so that $t_{k-1} < r \leqslant t_k$. The number of disjoint cycles in $\mathscr{L}_r$ is given by*

$$\frac{1}{k} \sum_{d \mid (u, k)} \phi(d) \binom{k/d}{u/d},$$

*where $(u, k)$ is the greatest common divisor of $u$ and $k$ and $\phi(d)$ is the Euler $\phi$-function, i.e., the number of positive integers at most $d$ and relatively prime to $d$.*

*Proof.* If $P = B(k-1) + \Delta \in \mathscr{L}_r \cap \mathscr{C}^{(k)}$, then $\beta^k(P) = P$ and so the period of $P \equiv \text{minimum}\{d > 0\colon \beta^d(P) = P\}$ is a divisor of $k$. Now let $d \mid k$ and see what it means for $\beta^d(P) = P$, or equivalently $\tilde{\beta}^d(\Delta) = \Delta$. If $A = \{i\colon \Delta_i = 1\}$ is the set of positions occupied by 1's on the stair, then after $d$ steps the list of 0's and 1's on the stair must look the same as the original list. This means that $A$ is the union of $k/d$ copies of $A \cap \{1, \dots, d\}$ each moved $d$ steps more to the right. Because $u$ is the cardinality of $A$ this requires $(k/d) \mid u$, i.e., $k \mid du$. Conversely, if $d \mid k$ and $k \mid du$, then such subsets exist and are determined by an arbitrary initial choice of $A \cap \{1, \dots, d\}$ as a subset of size $du/k$. So if we define

$$F(d) = \text{the cardinality of } \{P \in \mathscr{L}_r \cap \mathscr{C}\colon \beta^d(P) = P\},$$

then we have proved the formula:

$$F(d) = \begin{cases} \dbinom{d}{du/k}, & \text{if } d \mid k \text{ and } k \mid du, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

On the other hand, if we define

$$f(s) = \text{the cardinality of } \{P \in \mathscr{L}_r \cap \mathscr{C}\colon \text{period of } P = s\},$$

then $\beta^d(P) = P \Leftrightarrow$ period $P|d$ shows

$$F(d) = \sum_{s|d} f(s). \tag{27}$$

Finally, if $N$ is the number of disjoint cycles, then because each cycle of length $s$ contains $s$ elements of period $s$,

$$N = \sum_{s|k} f(s)/s. \tag{28}$$

Equation (27) implies that $F$ and $f$ are related by the Möbius inversion formula (see Niven and Zuckerman [**4**, p. 87]):

$$f(s) = \sum_{d|s} \mu(s/d) F(d).$$

Substituting in (28) we get:

$$\begin{aligned} N &= \sum_{s|k} \sum_{d|s} \mu(s/d)(1/s) F(d) \\ &= \sum_{d|k} \left[ \sum \{ \mu(s/d)(1/s) : d|s \text{ and } s|k \} \right] F(d). \end{aligned}$$

Let $z = s/d$; then $s|k$ if and only if $z|(k/d)$. Also, $1/s = (k/d)/z \times (1/k)$. Hence, the sum in brackets is

$$\frac{1}{k} \sum_{z|(k/d)} \mu(z)(k/d)/z.$$

By the Möbius inversion formula again, this is $\phi(k/d)/k$ (see Niven and Zuckerman [**4**, p. 88]). Letting $d' = k/d$ so that $d = k/d'$, we can substitute back into the equation above for $N$ to get:

$$N = \frac{1}{k} \sum_{d'|k} \phi(d') F(k/d').$$

Now $F(d) = F(k/d')$ is zero unless $k|du$, i.e., $d'|u$ and in the latter case $F(d) = \binom{k/d'}{u/d'}$. So the nonzero terms occur when $d'|k$ and $d'|u$, i.e., $d'|(u, k)$. QED

REMARK. If $u$ and $k$ are relatively prime (e.g., if $k$ is prime), then $d|(u, k)$ only when $d = 1$ and the formula reduces to $\binom{k}{u}/k$ as expected since all cycles have period equal to $k$.

As our story nears its end we feel the urge to leave some problems for the reader. So we introduce a new game called Austrian solitaire. It proceeds as follows:

Start by laying out stacks from a deck, but now all stacks have size $\leqslant L$, where $L$ is some fixed whole number. One special stack is reserved on the side. It is called the *bank*.

A round of play consists of two steps. First, remove one card from each ordinary stack and put it in the bank. Now from the bank lay out new stacks of size exactly equal to $L$, continuing until the size of the bank is $< L$ (including the possibility of exhausting the bank).

As with Bulgarian solitaire, the deck size is preserved and the problem is to describe the cycles. In particular, prove or disprove the following:

CONJECTURE. *For any fixed deck size and fixed L there is a unique cycle.*

We have only limited results which we would be happy to share with any reader who wants to play.

The idea for the game arose when we had returned to our usual pursuits and were reading a

discussion of the so-called Austrian school of capital theory. The economic interpretation shows that anything can become applied mathematics and also accounts for the name.

Think of the ordinary stacks as machines. Each machine has, when new, a life of exactly $L$ years. The size of a stack is the number of productive years left for a particular machine. Each year it ages one year (and so one card is removed from the stack). For each machine on line the company deposits $1/L$ of its cost into the bank as a sinking fund. Then it buys as many new machines as it can afford, and the remaining funds are left in the bank until next year.

Once these problems are solved the reader will be able to say, when asked about business cycles, "Why, it's all in the cards!"

### References

1. J. Brandt, Cycles of partitions, Proc. Amer. Math. Soc., 85 (1982) 483–486.
2. M. Gardner, Mathematical games, Scientific American, 249 (1983) 12–21.
3. R. May, Patterns of Species Abundance and Diversity, in Ecology and Evolution of Communities (Cody and Diamond, eds.), Harvard Univ. Press, 1975.
4. I. Niven and H. Zuckerman, An Introduction to the Theory of Numbers, Wiley, New York, 1962.

# SOME EXAMPLES OF COMBINATORIAL AVERAGING*

HERBERT S. WILF
*Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104*

Many important combinatorial problems, including some very hard ones, concern the determination of the extreme values of functions that are defined on graphs, sets, algorithms, etc. In some of these cases it turns out to be useful to try to learn more about the *distribution* of values of the functions that are involved. The maximum value, for instance, is one aspect of that distribution, but so are the average, the mean square, etc. Therefore one reason for studying combinatorial averaging is to shed some light on optimization problems.

Another motivation for studying discrete averages comes from the analysis of algorithms. In some computational problems the most natural measure of the running time might be the average time, averaged over a suitable collection of possible inputs, just as in many other computational problems we are interested more in the 'worst case', i.e., the maximum possible running time for the given set of inputs. In such situations the study of averages would be an end in itself rather than an aid to the solution of an extremum problem.

Whatever the reasons might be for looking at combinatorial averaging, the problems present great variety and beauty, along with some possibly very surprising outcomes. In this article I'd like to present a small sample of the varied fare of the subject.

My thanks go to Drs. E. Bender and D. E. Knuth for a number of helpful suggestions.

**1. A warmup: triangles in graphs.** In this section we will do quite an easy averaging problem on graphs. We will do it in two different ways, to illustrate the slightly different approaches of the *combinatorial model* and the *independent edge appearance model* of probabilistic graph theory.

Consider a graph $G$ of $n$ vertices, the vertices being labelled $\{1,2,\dots,n\}$. There are $2^{\binom{n}{2}}$ such graphs, since each of the $\binom{n}{2}$ possible pairs of vertices might or might not be joined by an edge of the graph.

For each such graph $G$, let $T(G)$ be its number of triangles, i.e., the number of triples $i, j, k$ of vertices such that $(i,j),(i,k),(j,k)$ are all edges of $G$.

*Presented, by invitation, at the Annual Meeting of the Association in Louisville, Kentucky, January, 1984.

The maximum of $T(G)$, over all graphs $G$ of $n$ vertices, is clearly $\binom{n}{3}$; the minimum is 0. What is the *average* number of triangles in graphs of $n$ vertices?

First, the combinatorial model. We look at the number of triangles in each graph, add them up over all graphs, and divide by the number of graphs, to get

$$\overline{T}(n) = \frac{1}{2^{\binom{n}{2}}} \sum_G T(G) \tag{1}$$

in which the sum is over all $2^{\binom{n}{2}}$ graphs $G$ of $n$ vertices.

(*We have just carried out the first step in a rather general three step recipe for computing the average number of substructures within a structure, which is to express that average as a suitable sum over the structures.*)

Now we are going to introduce yet another summation sign, involving a function whose values are only 0's and 1's. $T(G)$, in this example, is the sum, over all $\binom{n}{3}$ triples of vertices, of a function whose value is 1 if a triangle of $G$ lives at that triple and 0 otherwise. Hence define $\chi(i, j, k; G) = 1$ if $(i, j), (j, k), (i, k)$ are all edges of $G$, $\chi = 0$ otherwise. Then we can express $T(G)$ as

$$T(G) = \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} \chi(i, j, k; G), \tag{2}$$

and so from (1),

$$\overline{T}(n) = 2^{-\binom{n}{2}} \sum_G \sum_{i,j,k} \chi(i, j, k; G). \tag{3}$$

(*Now we've done the second step, in such problems, in which we look for an 'indicator function' $\chi$, whose values are just* 0'*s and* 1'*s, that allows us to express the average as a double sum of* 0'*s and* 1'*s.*)

This is the time to interchange the summation signs, which gives

$$\overline{T}(n) = 2^{-\binom{n}{2}} \sum_{i,j,k} \left\{ \sum_G \chi(i, j, k; G) \right\}, \tag{4}$$

and we now face quite a different kind of sum inside the curly braces. There we find, for a *fixed* triple $i, j, k$, the number of graphs $G$ that have a triangle at $i, j, k$.

Well then, how many graphs $G$ have a triangle at vertices $1, 2, 3$? If there is a triangle there then 3 of the $\binom{n}{2}$ possible edges of the graph $G$ have been determined: they exist. There remain just $\binom{n}{2} - 3$ edges of $G$ that might or might not exist, and so there are exactly $2^{\binom{n}{2}-3}$ such graphs $G$.

(*And that was the third step: to interchange the order of the summations, and to try, as above, to deal with the innermost sum.*)

For each of the $\binom{n}{3}$ triples $i, j, k$, then, the inner sum in (4) has the value $2^{\binom{n}{2}-3}$, and we find at once that an average graph of $n$ vertices has

$$\overline{T}(n) = \frac{1}{8}\binom{n}{3} \tag{5}$$

triangles.

The answer (5) for the average number of triangles just begs for a 'one line proof', because on probabilistic grounds it is obvious, and that brings us, as promised, to the independent edge appearance model, pioneered by Erdős and Rényi **[12]**. In this model we have a fixed edge probability $p$, $0 < p < 1$. An event in our probability space is obtained by tossing a (biassed) coin $\binom{n}{2}$ times, once for each possible edge in a graph of $n$ vertices, where the two possible outcomes

of each toss are called 'edge' and 'no edge', respectively, and in which the probability of 'edge' is $p$. Whenever the outcome is 'edge' we draw an edge, else we do not, and this gives us the graph $G$ that was sampled by the repeated tosses of the loaded coin.

Now consider the number of triangles. If $i < j < k$ are fixed vertices, the probability that all three edges $(i, j), (i, k), (j, k)$ occur is $p^3$, and immediately we see that the average number of triangles is $p^3\binom{n}{3}$, in agreement with (5) when $p = \frac{1}{2}$. The independent edge appearance model is usually easier to handle, as this example shows. There may be small differences in the answers that we get with the two models. To do the combinatorial model with 'edge probability $p$' we would sum over just those graphs of $n$ vertices that have exactly $\left\lfloor p\binom{n}{2}\right\rfloor$ edges, instead of summing over all $2^{\binom{n}{2}}$ of them. The average number of triangles will be slightly different from $p^3\binom{n}{3}$ (work it out!), but the differences will not be important for most applications.

EXERCISE 1. How many complete graphs of $k$ vertices, on the average, does a graph of $n$ vertices $(n \geqslant k)$ contain? (Use either model.)

The familiar inequality of Tschebycheff, from probability theory (e.g., **[14]**, p. 233) gives information about a full probability distribution from information about just the first two moments of that distribution. More precisely, if we know the mean $\mu$ and the standard deviation $\sigma$ of the distribution of a certain random variable $x$, then the inequality tells us that $x$ is unlikely to be found very far away from its mean value, in fact that for every positive $t$,

$$\text{Prob}\{|x - \mu| > t\sigma\} < \frac{1}{t^2}. \tag{6}$$

In view of Tschebycheff's inequality, there is a special value in finding the first *two* moments of an unknown distribution, since we then get at least some information about the shape of the whole curve. If you would like to see this principle in action, then try

EXERCISE 2. What is the standard deviation of the number of triangles in graphs of $n$ vertices? What does Tschebycheff's inequality tell us about the probability that a graph has a large number of triangles? (Use either model.)

**2. The fixed points of permutations.** Here's another example of the same technique. If $n$ is fixed, what is the average number of fixed points that permutations of $n$ letters have, and what is the standard deviation of that number? Let $S_n$ be the set of all permutations of $n$ letters, and let $\sigma \in S_n$. By $f(\sigma)$ we mean the number of fixed points of $\sigma$, and let $\bar{f}(n)$ be the average that we seek. Then (*step* 1)

$$\bar{f}(n) = \frac{1}{n!}\sum_{\sigma \in S_n} f(\sigma). \tag{7}$$

Now (*this is step* 2 *of the recipe*) we introduce the indicator function $\chi(i; \sigma) = 1$ if $\sigma(i) = i$; $= 0$ otherwise, where $1 \leqslant i \leqslant n$. Then (7) expands, just as in (3), this time to

$$\bar{f}(n) = \frac{1}{n!}\sum_{\sigma \in S_n}\sum_{i=1}^{n}\chi(i; \sigma).$$

Once again (*here comes step* 3) we invoke the commutative law of addition to interchange the summation signs and find that

$$\bar{f}(n) = \frac{1}{n!}\sum_{i=1}^{n}\left\{\sum_{\sigma \in S_n}\chi(i; \sigma)\right\}. \tag{8}$$

The inner sum is the number of permutations that have fixed point $i$, and that is clearly $(n - 1)!$ Hence

$$\bar{f}(n) = \frac{1}{n!} \sum_{i=1}^{n} (n-1)! = 1. \tag{9}$$

Since you probably already knew the answer (9) that average permutations have 1 fixed point, let's try the second moment of the distribution, i.e., the *mean square* of the number of fixed points. This is

$$Q(n) = \frac{1}{n!} \sum_{\sigma \in S_n} f(\sigma)^2. \tag{10}$$

Again the strategy will be to boil the sum down to a '0 or 1' summand, interchange the order of summations, and hope. We find that

$$Q(n) = \frac{1}{n!} \sum_{\sigma \in S_n} \left\{ \sum_{i=1}^{n} \chi(i;\sigma) \right\}^2. \tag{11}$$

The square of the sum is handled by writing down the sum twice, using two different indices of summation. In other words we use the fact that

$$\left\{ \sum_{i=1}^{n} \alpha_i \right\}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j$$

and then (11) becomes

$$Q(n) = \frac{1}{n!} \sum_{\sigma \in S_n} \sum_{i=1}^{n} \sum_{j=1}^{n} \chi(i;\sigma)\chi(j;\sigma). \tag{12}$$

If we interchange the summations in (12),

$$Q(n) = \frac{1}{n!} \sum_{i,j=1}^{n} \left\{ \sum_{\sigma \in S_n} \chi(i;\sigma)\chi(j;\sigma) \right\}. \tag{13}$$

What sort of creature now inhabits the curly braces? For a fixed pair $i, j$ we are counting the permutations that fix *both* $i$ and $j$. The number of these is $(n-1)!$ if $i = j$ and $(n-2)!$ if $i \neq j$. Therefore, if $n > 1$ we have

$$\begin{aligned} Q(n) &= \frac{1}{n!} \left\{ \sum_{i=1}^{n} (n-1)! + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} (n-2)! \right\} \\ &= 2. \end{aligned} \tag{14}$$

EXERCISE 3. The mean is 1; the mean square is 2; what is the mean cube? Is it an integer?

The standard deviation of the number of fixed points is therefore 1. By Tschebycheff's inequality (6), the probability that a permutation has more than $t + 1$ fixed points is less than $1/t^2$.

**3. In search of the biggest determinant.** If $M$ and $n$ are fixed, and if we know that no entry of a certain $n \times n$ matrix $A$ exceeds $M$, in absolute value, then how big can $|\det A|$ be?

This is the famous (and unsolved) determinant problem of Hadamard. We might as well standardize things by taking $M = 1$. Then it's easy to check that the maximum of $|\det A|$ is attained on some matrix whose entries are all $\pm 1$'s. Since there are a mere $2^{n^2}$ such matrices, the problem is now finite. Since $2^{n^2}$ is a very healthy number when $n$ is only modestly large, an exhaustive search-by-computer for the maximum is already out of the question for $n \geqslant 6$, or thereabouts.

Hadamard observed that since the absolute value of the determinant measures the volume of the parallelepiped spanned by the row vectors, and since, if the lengths of the row vectors are fixed ($\sqrt{n}$ in this case) that volume is largest when the row vectors are pairwise orthogonal, we have

THE HADAMARD DETERMINANT THEOREM. *The absolute value of the determinant of a matrix is at most equal to the product of the lengths of its row vectors. It is exactly that large if and only if the matrix is orthogonal.*

In the present case, it follows first that the value of the determinant of an $n \times n$ matrix of $\pm 1$'s is at most $n^{n/2}$, and second, that it can be that large if and only if there is such a matrix whose rows are pairwise orthogonal. An $n \times n$ orthogonal matrix of $\pm 1$'s is called an *Hadamard matrix*. It is easy to see that $n$ must be a multiple of 4 if an Hadamard matrix of size $n$ exists, but it is not known which multiples of 4 will work, though a good deal of research has been done on the question.

Since the *maximum* determinant is so hard to corner, what about some averages instead? The average determinant of all $\pm 1$ matrices of size $n$ is obviously 0, so that won't be of much help. The mean square looks more promising, and it is.

The following amazing averaging calculation was shown to me by Mark Kac, many years ago. It is amazing to me because the answer turns out to be so simple, just one line after the whole thing starts to look hopelessly complicated.

We let $D_n^2$ denote the average of the squares of the determinants of all $n \times n$ matrices (all $2^{n^2}$ of them) whose entries are $\pm 1$'s. Then directly from the definition of the determinant we have

$$\begin{aligned} D_n^2 &= \frac{1}{2^{n^2}} \sum_A (\det A)^2 \\ &= 2^{-n^2} \sum_A \left\{ \sum_\pi \operatorname{sign}(\pi)\, a_{1,i_1} a_{2,i_2} \cdots a_{n,i_n} \right\}^2, \end{aligned}$$

in which the outer sum is over all matrices of this kind, and the inner sum is over the $n!$ permutations $(i_1, \ldots, i_n)$ of $n$ letters.

Just after equation (11) I said that when the square of a sum is encountered, the thing to do is to write out the sum twice, using two different names for the summation variable. If we (swallow hard and) follow that advice here, we find that

$$D_n^2 = 2^{-n^2} \sum_A \sum_\pi \sum_\rho \operatorname{sign}(\pi) \operatorname{sign}(\rho) \{ a_{1,i_1} a_{1,j_1} a_{2,i_2} a_{2,j_2} \cdots a_{n,i_n} a_{n,j_n} \}, \tag{15}$$

where $\pi$ and $\rho$ vary independently over the permutations of $n$ letters. After interchanging the order of the summations in (15) we get

$$D_n^2 = 2^{-n^2} \sum_\pi \sum_\rho \operatorname{sign}(\pi) \operatorname{sign}(\rho) \left\{ \sum_A a_{1,i_1} a_{1,j_1} a_{2,i_2} a_{2,j_2} \cdots a_{n,i_n} a_{n,j_n} \right\}. \tag{16}$$

Things are now as dark as they will ever get, and dawn is rapidly approaching: Consider the sum in the inner curly brace. In that sum we have a *fixed pair* $\pi, \rho$ of permutations, and we are summing over all of the $\pm 1$ matrices. That means that '$\Sigma_A$' really stands for $n^2$ summation signs, each one over a single summation variable, in which that variable takes just two values: $+1$ and $-1$. If $n = 2$, for instance, the inner sum looks like this

$$\sum_{a_{1,1} = \pm 1} \sum_{a_{1,2} = \pm 1} \sum_{a_{2,1} = \pm 1} \sum_{a_{2,2} = \pm 1} a_{1,i_1} a_{1,j_1} a_{2,i_2} a_{2,j_2}. \tag{17}$$

This unpleasant looking sum has one saving grace: it loves to be 0. Suppose, for instance, that $i_1 \neq j_1$; say $i_1 = 1$, $j_1 = 2$. Then the sum (17) is 0 because it has as a factor

$$\sum_{a_{1,1}=\pm 1} a_{1,1} = 0. \tag{18}$$

*The only way that* (17) *can fail to vanish is if* $i_1 = j_1$ *and* $i_2 = j_2$.

Likewise, in (16), the inner sum vanishes unless $\pi = \rho$, and suddenly everything is easy. We have

$$\begin{aligned} D_n^2 &= 2^{-n^2} \sum_{\pi} \{\operatorname{sign}(\pi)\}^2 2^{n^2} \\ &= n!. \end{aligned} \tag{19}$$

The mean square determinant of all of these matrices is $n!$. Their root-mean-square is $D_n = \sqrt{n!}$, and when $n$ is large this is

$$\sim (2\pi n)^{1/4} \frac{n^{n/2}}{e^{n/2}}, \tag{20}$$

by Stirling's formula. We observe that the maximum possible determinant, by Hadamard's theorem, is $n^{n/2}$, and the root-mean-square isn't much lower than that.

Since the maximum cannot be less than the root-mean-square, we have *proved* the following

THEOREM. *There exists an* $n \times n$ *matrix of* $\pm 1$*'s whose determinant is greater than* $\sqrt{n!}$.

It is characteristic of averaging methods that while we learn that such a matrix exists, we haven't a clue as to how to construct one efficiently.

Those who enjoy problems about permutations (*permutophiles*?) might like to try

EXERCISE 4. Show that the root-mean-square value of the determinants of all $n \times n$ matrices of 0's and 1's is $(n+1)!/4^n$. Show that the average of these determinants is 0, without doing any sums!

**4. The complexity of computations.** The averaging methods that we have described can be used to determine the average complexities of certain algorithms. We will illustrate this with the problem of finding the maximum independent set of a graph. In this problem, a graph $G$ is given, and we are looking for a set $S$ of vertices of $G$, as large a set as possible, with the property that no two vertices of $S$ are joined by an edge of $G$.

If this problem seems simple, it isn't. It is a member of the notorious clan of '*NP*-complete' problems, which for the present purposes means that the question is seemingly, but not at present provably, intractable from a computational point of view. That is, all known methods for solving it computationally require an amount of time that is exponentially large as a function of the length of the string of input data.

Anyway, let's analyze the *average* time complexity of the most obvious method of doing this problem: exhaustive search. In that method we simply look at each and every independent set of vertices in the graph, and find the largest one. It is easy to march from one independent set to its successor (how?) quickly, so the amount of labor involved in this procedure is fairly measured by the *number* of independent sets in the graph.

Now, in the worst case, a graph of $n$ vertices can have quite a few independent sets, namely as many as $2^n$ of them (which graph has that many?). Our question, though, is not about the worst case, but the average case: on the average, how many independent sets does a graph of $n$ vertices have?

Let $I(G)$ be the number of independent sets of vertices in a graph $G$, and let $\bar{I}_n$ be the average of $I(G)$ over all graphs $G$ of $n$ vertices. We use the independent edge appearance model to find $\bar{I}_n$. If $T$ is a fixed set of exactly $t$ vertices, then $T$ will be an independent set if and only if all $\binom{t}{2}$ of the possible edges that might join pairs in $T$ in fact fail to materialize when the 'coins are tossed.' If the edge probability is $p$, then the probability that $T$ turns out to be an independent set

is evidently $(1-p)^{\binom{t}{2}}$. Hence the average number of independent sets of vertices in graphs of $n$ vertices is

$$\bar{I}_n = \sum_{t=0}^{n} \binom{n}{t}(1-p)^{\binom{t}{2}}, \tag{21}$$

since exactly $\binom{n}{t}$ subsets $T$ of $[1,\ldots,n]$ have cardinality $t$.

How bad is the news in (21)? That is, is (21) a large amount of work for an algorithm to do, on the average, or not so large?

First let's find out which of the $n$ terms in the sum (21) is the biggest one. Then $n$ times that one term will be an upper estimate for the whole sum, and therefore for the average labor involved in the algorithm that we are analyzing. To keep things simple we choose $p = \frac{1}{2}$.

We observe that the terms in (21) increase in size for a while (i.e., for small $t$) and then they decrease. To find the largest term we therefore take the ratio of the $t$th term to the $(t-1)$st, set the ratio equal to 1 and solve for $t$. After a bit of manipulation, this ratio of consecutive terms works out to

$$\frac{(n-t+1)}{t2^{t-1}}.$$

The ratio is therefore equal to 1 when $t2^{t-1} = n - t + 1$. Clearly this value of $t$ is not very far from $\log_2 n$, and if $t_n$ denotes the value of $t$ in question, then it is easy to check that

$$\log_2 n - \log_2 \log_2 n < t_n < \log_2 n.$$

After another brief estimation, this time an exercise in Stirling's formula, the size of the largest term is seen to be

$$< \exp\left\{\left(\frac{1}{2} + \varepsilon\right)(\log_2 n)^2\right\}. \tag{22}$$

Hence the average amount of computation, being at most $n$ times the quantity shown in (22), also satisfies (22).

This amount of labor, in a computation, grows more rapidly than any polynomial in $n$, but less rapidly than an exponential function $c^n$ $(c > 1)$. It is therefore referred to as being of *moderately exponential* growth. To summarize, then, the *worst case* for this algorithm clearly involves an exponentially-in-$n$ growing amount of computational labor, but the *average* is a bit tamer.

EXERCISE 5. Use the independent edge appearance model with edge probability $p$ to find a formula, corresponding to (21), for the average number of maximal independent sets in graphs of $n$ vertices (an independent set is maximal if it is not a proper subset of an independent set).

RESEARCH PROBLEM. Find the asymptotic behavior, for $n \to \infty$, of the number of independent sets and of the number of maximal independent sets of vertices in graphs of $n$ vertices. Do this for edge probabilities throughout the range from $p \sim C/n$ to $p \sim C$.

**5. Two surprises.** The example that we worked out in the previous section, of the average number of independent sets in graphs, illustrates a technique and an approach to complexity analysis of algorithms, but there are other applications of the technique that give more surprising results. We describe two such recent results in this section, without proofs.

Consider the question of deciding if the vertices of a given graph $G$ can be properly colored in 3 colors (i.e., so that adjacent vertices always have different colors). This is the graph 3-coloring problem, and it is $NP$-complete.

Suppose we try to find a 3-coloring by backtracking:

(a) Assign a color to vertex 1.
(b) In general, if vertices $1, 2, \ldots, k$ have been properly colored, then assign a color to vertex $k+1$ that doesn't conflict with the colors that have already been assigned to lower-num-

bered vertices (if such a color exists), increase $k$ by 1, and go to (b); if no such color exists then go to (c).

(c) Recolor vertex $k$ in the next free color for that vertex (if one exists), and go to (b); else, if there are no remaining free colors for vertex $k$, then reduce $k$ by 1 and go to (c).

(d) Halt if a proper coloring of $G$ is found (success) or if the algorithm ever attempts to re-color vertex 1 (failure).

Backtracking *might* take a very long time. For instance, if the four highest numbered vertices of $G$ are all connected to each other by edges, and no other edges exist in $G$, then $G$ cannot be 3-colored, but the algorithm will take nearly $3^n$ steps to find that out, if $G$ has $n$ vertices.

How many steps *on the average* does backtracking take? If we average the backtracking time over all graphs on $n$ vertices we find that the average is approximately (guess which one!):

(a) $3^n$
(b) $e^{\sqrt{n}}$
(c) $cn^2$
(d) 197
(e) none of the above.

The correct answer is 197 (see [**9**] for the proof and [**10**] for generalizations and refinements) and that fact, hopefully, is the first of the two surprises in this section.

The 'reason' for this result is that for most graphs the answer is 'no' (i.e., $G$ cannot be 3-colored), and indeed there is a complete graph of 4 vertices in almost all graphs that usually stops the algorithm quickly, with a 'no' answer.

The 'reason', however, isn't in itself sufficient. In another important computational problem, the graph isomorphism problem, most pairs of random graphs are not isomorphic, and almost always we can find that out quickly. However, the few pairs that are isomorphic generate huge computations, and so the averages still turn out to be exponentially large, at least for some of the algorithms that have been proposed, despite the heavy preponderance of 'no' answers.

The second surprise comes from the bin-packing problem. In that problem we are given $n$ real numbers $\xi_1, \dots, \xi_n$, where $0 < \xi_i < 1 \quad (i = 1, n)$. We are asked if we can 'pack' these numbers into a given number of 'bins' of unit length (a set of numbers is *packed* into a unit bin if the sum of those numbers is at most 1). The problem is $NP$-complete.

One well-known algorithm for bin-packing is called 'First-Fit-Decreasing' (FFD). In FFD we first sort the given $\xi_1, \dots, \xi_n$ into nonincreasing order of size. Then, for each $\xi$ in turn, taken in the sorted order, we do 'put $\xi$ into the first bin that it fits into.'

This algorithm, of course, doesn't always find the optimum packing, i.e., the packing into the smallest number of bins. In the aptly titled [**13**], the *average* behavior of FFD was studied, as follows. Fix a number $u$, $0 \leqslant u \leqslant \frac{1}{2}$. Suppose that $\xi_1, \dots, \xi_n$ are chosen independently uniformly at random from $[0, u]$, and then suppose that FFD is used to pack them into bins. A result of [**13**] is that there is a constant $C > 0$ such that the average excess of bins used by FFD over the minimum possible number of bins is $< C$, averaged over all inputs $\xi_1, \dots, \xi_n$ and for all $n$. There is experimental evidence that suggests that FFD *gives optimal packing about* 75 *percent of the time*, but this has not been proved. This result is more striking because there is no feeling that 'cheating' is going on because most answers are 'no', or whatever.

**6. The hierarchy of average powers.** We have already referred to the fact that the maximum of a finite set of positive real numbers can never be less than the root-mean-square value of the same set of numbers. This is a special case of a much more general pecking order that holds between various kinds of means of the same set of numbers. In this section we will state and prove a rather general form of this result, and in the next section there will be a combinatorial application of the idea, in a somewhat surprising way. The 'pecking order' theory is quite elementary, but somehow seems to be not as widely known as it deserves to be.

Let $a_1, \ldots, a_n$ be a fixed set of positive reals. If $r$ is a real number, then by the $r$th *power mean* of the $a$'s we mean the number

$$\mu(r) = \left\{ \frac{a_1^r + a_2^r + \cdots + a_n^r}{n} \right\}^{1/r}. \tag{23}$$

The power means have a number of remarkable and useful properties. First, consider a few special values of $r$. The case $r = 1$ is perhaps the most obvious: $\mu(1)$ is the well-known *arithmetic mean* of the given numbers. Another easy special value is $r = 2$. Inspection of (23) shows that $\mu(2)$ is exactly the root-mean-square (rms) of the given set of $a$'s.

Slightly less obvious is the value '$r = \infty$', or, more precisely, the value of $\lim_{r \to \infty} \mu(r)$. We leave it as an exercise for the reader to verify that

$$\mu(\infty) = \max_{1 \leqslant i \leqslant n} a_i.$$

That $\mu(-\infty) = \min a$ is then unsurprising, and the special value

$$\mu(-1) = \frac{n}{\left\{ \dfrac{1}{a_1} + \dfrac{1}{a_2} + \cdots + \dfrac{1}{a_n} \right\}}$$

is the familiar *harmonic mean* of the given set of $a$'s. The best one has been saved for last, though. What is $\mu(0)$ (the limit, that is)? The answer is in

EXERCISE 6. Prove that $\mu(0)$ is the geometric mean of the $a$'s, i.e.,

$$\mu(0) = \{a_1 a_2 \cdots a_n\}^{1/n}.$$

The exercise above requires only a small application of l'Hospital's rule, but it is highly recommended to those who haven't tried it.

Let's use special names for the six special kinds of means of a given set of positive numbers that we have just discussed. We will write, respectively, $\mathscr{L}(a)$, $\mathscr{H}(a)$, $\mathscr{G}(a)$, $\mathscr{A}(a)$, $\mathscr{R}(a)$ and $\mathscr{U}(a)$ for the minimum, the harmonic mean, the geometric mean, the arithmetic mean, the root mean-square, and the maximum, respectively, of a set $a$ of positive numbers.

The fact that the power mean function $\mu(r)$ takes all of these interesting values, at $r = -\infty, -1, 0, 1, 2, +\infty$, lends a special impact to the main theorem of the subject, which is

THE POWER-MEAN THEOREM. *If $a_1, \ldots, a_n$ is a set of positive real numbers, then $\mu(r)$ is a strictly increasing function of the real variable $r$ unless $a_1 = a_2 = \cdots = a_n$, in which case $\mu$ is constant for all real $r$.*

This theorem shows that for every nonconstant set of $a$ we have, in particular,

$$\mathscr{L}(a) < \mathscr{H}(a) < \mathscr{G}(a) < \mathscr{A}(a) < \mathscr{R}(a) < \mathscr{U}(a).$$

The guaranteed lineup of the various kinds of averages shows that, for instance, if we want to estimate the maximum, and that's too hard, we might try the root-mean-square next, because it is 'closest' to the maximum, then the arithmetic mean, etc.

The other feature of the theorem that merits comment is the strong grip that we have on the case of equality: *if there exist $r_1$ and $r_2$, such that $r_1 \neq r_2$ and $\mu(r_1) = \mu(r_2)$, then we can be sure that all $n$ of the given set $a_1, \ldots, a_n$ of numbers are equal to each other.* This strong conclusion that follows from the equality of any two of the power means has a number of applications, one of which we will see in Section 6 below.

Now, as to the proof of the power-mean theorem, we begin with

LEMMA 1. *Let $\phi(t)$ be twice differentiable, with $\phi''(t) > 0$, and let $w_1, \ldots, w_n$ be nonnegative numbers whose sum is* 1. *Then*

$$\phi(w_1 t_1 + w_2 t_2 + \cdots + w_n t_n) \leqslant w_1 \phi(t_1) + w_2 \phi(t_2) + \cdots + w_n \phi(t_n) \tag{24}$$

*and the sign of equality holds if and only if the $t_i$ are all equal.*

*Proof.* If we put $A = w_1 t_1 + w_2 t_2 + \cdots + w_n t_n$ then $\forall i = 1, n$, there exists a $\xi_i$ such that

$$\phi(t_i) = \phi(A) + (t_i - A)\phi'(A) + \frac{(t_i - A)^2}{2}\phi''(\xi_i).$$

If we multiply by $w_i$ and sum over $i$ the result follows immediately. ∎

If we set $\phi(t) = -\log t$ and replace the $t_i$ by $1/t_i$ in (24), we get

$$w_1 \log t_1 + \cdots + w_n \log t_n > -\log\left\{\frac{w_1}{t_1} + \cdots + \frac{w_n}{t_n}\right\} \tag{25}$$

in which the strict inequality holds if and only if the $t_i$ are not all equal.

To prove the theorem we differentiate the power mean function $\mu(r)$, directly from its definition (23), and then write

$$w_i = \frac{a_i^r}{a_1^r + \cdots + a_n^r} \qquad (i = 1, n). \tag{26}$$

The result is that

$$r^2 \frac{\mu'(r)}{\mu(r)} = w_1 \log a_1^r + \cdots + w_n \log a_n^r - \log\left\{\frac{a_1^r + \cdots + a_n^r}{n}\right\}. \tag{27}$$

Since $\Sigma w_i = 1$, the inequality in (25) yields at once the fact that the right member of (27) is strictly positive unless all of the $a$ are equal, completing the proof. ∎

For a fixed set of $a_1, \ldots, a_n$ a typical graph of the function $\mu(r)$ vs. $r$ is as shown in Fig. 1.

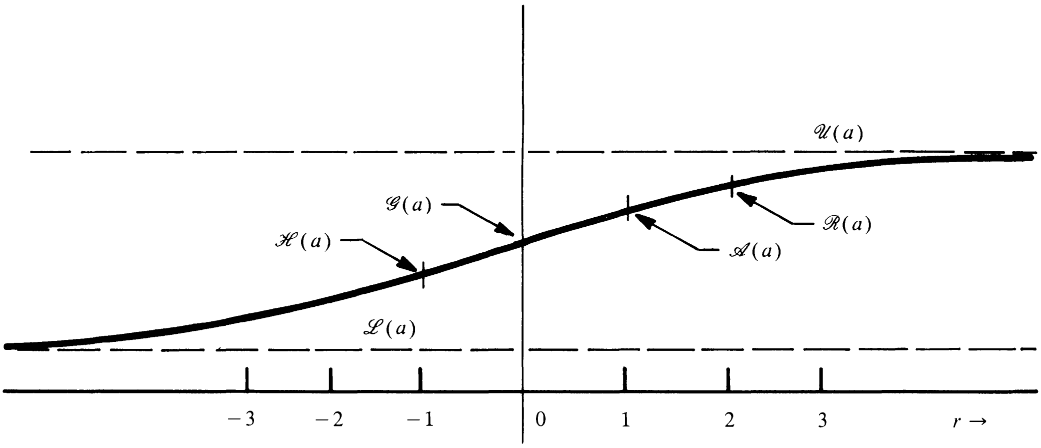



FIG. 1

**7. Cyclic designs.** In this section we will give a small application of the power-sum theorem that was just proved. The theorem will be applied in the following way (which may seem suspicious because we will get $n$ pieces of information in exchange for just two):

(1) We will want to show that a certain collection of $n$ numbers are all the same.
(2) We will compute their geometric mean $\mathscr{G}$.
(3) We will compute their root-mean-square $\mathscr{R}$.
(4) We will note that $\mathscr{G} = \mathscr{R}$, and therefore we will conclude that all $n$ of the numbers are equal.

The combinatorial question is this: for which $n$ does there exist an $n \times n$ *cyclic* orthogonal matrix of $\pm 1$'s? We are therefore looking for an Hadamard matrix, as described in Section 3 above, with the additional restriction that the matrix must be cyclic.

We have first the following well-known evaluation of a general cyclic determinant:

$$(28)\qquad \det\begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \cdots & a_0 \end{pmatrix} = \prod_{\omega^n=1} f(\omega)$$

in which $f$ is the polynomial whose coefficients appear in the first row of the given matrix, i.e., $f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_{n-1}t^{n-1}$.

The proof of (28) is easy. Let $A(a_0, a_1, \ldots, a_{n-1})$ be the matrix that appears on the left, and write $C = A(0,1,0,\ldots,0)$. The eigenvalues of $C$ are the $n$th roots of unity $\omega$, and since $A(a_0, a_1, \ldots, a_{n-1}) = f(C)$, the numbers $f(\omega)$ are the eigenvalues of $A$, and their product is the determinant. ■

What we are going to prove next is the following

THEOREM. *There exists an $n \times n$ cyclic Hadamard matrix if and only if there is a polynomial $f$, of degree $n-1$, whose coefficients are all $\pm 1$'s, and whose values at the $n$th roots of unity are all equal, in absolute value.*

As advertised above, let us first compute the geometric mean of the $n$ numbers

$$(29)\qquad |f(\omega)| \qquad (\omega^n = 1).$$

Since the matrix is orthogonal, its determinant is, in absolute value, equal to the product of the lengths of its row vectors, i.e., to $n^{n/2}$, since each row has length $\sqrt{n}$. From the determinant formula (28), if we take the absolute value, and then the $n$th root of both sides, we find that *the geometric mean of the $n$ numbers shown in* (29) *is* $\sqrt{n}$.

We claim that *the root-mean-square value of these same $n$ numbers is also* $\sqrt{n}$. Indeed, the reader will have no difficulty verifying that this is true for every polynomial $f$ whose coefficients are $\pm 1$'s.

Since the reverse implication is clear, this concludes the proof of the theorem. We remark that the proof showed that the 'constant' value of the numbers in (29) is none other than $\sqrt{n}$. Since the integer $|f(1)|$ is among the values in (29), it too must be $\sqrt{n}$ and we have the

COROLLARY. *If there is a cyclic $n \times n$ Hadamard matrix, then $n$ is a perfect square.*

**8. Suggestions for further reading.** A number of properties of 'average trees' are in the lovely book [**1**] of Moon, such as the average number of leaves, the average and second moment of the number of spanning trees of a graph, the average valence of the vertex of largest valence, etc.

The book of Comtet [**2**] is highly recommended as a general reference in combinatorics, and his Chapter 7 contains a number of averaging problems and solutions.

I have borrowed the treatment of power sums from the classic work of Pólya-Szegö [**3**], in which Chapter 2 of Section 2 ('inequalities') is a rich source of information about different kinds of averages.

There is a proof of the Hadamard determinant theorem in Hardy, Littlewood, and Pólya [**4**]. An introduction to the combinatorial problems surrounding that theorem is in Ryser [**5**], whereas an advanced and complete-as-of-its-date account of the theory of Hadamard matrices can be found in [**6**].

Paul Erdős has made many important contributions to probabilistic combinatorics and graph theory. A fine survey of these is in Erdős and Spencer [**7**]. The three volumes of Knuth [**8**] are an excellent source of probabilistic combinatorics and algorithmic applications.

Another averaging result of unusual interest is that of Grimmett and McDiarmid [**11**] on the chromatic number of a graph.

**References**

1. J. W. Moon, Counting Labelled Trees, Canad. Math. Monographs, no. 1, Canadian Mathematical Congress, 1970.

2. L. Comtet, Advanced Combinatorics, D. Reidel, Boston, 1974.

3. G. Pólya and G. Szegö, Aufgaben und Lehrsätze aus der Analysis, Springer-Verlag, Berlin, 1954.

4. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge Univ. Press, 1959.

5. H. J. Ryser, Combinatorial Mathematics, Carus Mathematical Monographs, no. 14, Mathematical Association of America, Washington, DC, 1963.

6. W. D. Wallis, Anne Penfold Street, and Jennifer Seberry Wallis, Combinatorics: Room Squares, Sum-Free Sets, Hadamard Matrices, Lecture Notes in Mathematics, no. 292, Springer-Verlag, New York, 1972.

7. P. Erdős and J. Spencer, Probabilistic Methods in Combinatorics, Probability and Mathematical Statistics, vol. 17, Academic Press, New York, 1963.

8. D. E. Knuth, The Art of Computer Programming (3 vols.), Addison-Wesley, Reading, MA, 1968, 1969, 1973.

9. H. S. Wilf, Backtrack: An $O(1)$ expected time algorithm for the graph coloring problem, Information Processing Letters, 18 (1984) 119–122.

10. E. Bender and H. S. Wilf, A theoretical analysis of backtracking in the graph coloring problem, J. Algorithms, to appear.

11. G. R. Grimmett and C. J. H. McDiarmid, On coloring random graphs, Math. Proc. Cambridge Philos. Soc., 77 (1975) 313–324.

12. P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci., 5 (1960) 17–61.

13. J. L. Bentley, D. S. Johnson, F. T. Leighton, C. C. McGeoch, and L. A. McGeoch, Some unexpected expected behavior results for bin packing, 1984 (preprint).

14. W. Feller, An Introduction to Probability Theory and Its Applications, vol. 1, Wiley, New York, 1968.


# MS. MACHIAVELLI AND THE STABLE MATCHING PROBLEM


DAVID GALE

*Department of Mathematics, University of California, Berkeley, California 94720*

MARILDA SOTOMAYOR*

*Department of Mathematics, Pontificia Universidade Catolica do Rio de Janeiro, Brazil*


This paper is a sequel to a paper, "Machiavelli and the Gale-Shapley Algorithm," written by Dubins and Freedman [**1**] in 1981. That paper was, in turn, a sequel to "College Admissions and the Stability of Marriage," by Gale and Shapley [**2**], 1962. For the benefit of readers who may have missed the previous installments, a brief recapitulation is given in the following section.

**1. The Story So Far.** Paper [2] above was concerned with a situation in which there are two sets of "agents", such as students and universities or workers and employees, or women and men. For the sake of concreteness, we will describe the problem in terms of this last group. It is assumed that each man and woman has prepared a list (possibly empty) containing the names of

Marilda Sotomayor is an assistant professor of mathematics at the Catholic University of Rio de Janeiro, where she obtained her Ph.D. in mathematics. Her thesis, in the area of mathematical economics, was supervised by Jack Schechtman who headed an applied mathematics group at the Institute for Pure and Applied Mathematics in Rio. Schechtman has a Ph.D. in operations research from Berkeley, where he wrote a thesis in mathematical economics, supervised by David Gale. Gale has been writing about games and economics on and off since obtaining his Ph.D. in mathematics from Princeton at some uncertain date, believed to be after the middle Pleistocene era.

*Supported by CNPq (Conselho Nacional de Desenvolvimento Cientifico e Technologico), Brazil.

those people of the opposite sex that he or she would accept as a marriage partner and that the names appear on the list in order of preference. The lists are then submitted to a matchmaker who has the job of finding a suitable pairing of the participants. If we think of an unmatched person as being matched with him or herself, then the only requirement for the matching is that it should be free of *instability*; that is, there should not be any man $m$ and woman $w$ who are not matched with each other but prefer each other to their mates. The matchings are thus, "divorce-proof". The main result of [**2**] is an existence theorem which shows that for any sets $M$ and $W$ of men and women and any preference pattern, there is always at least one *stable matching*. The proof is constructive, giving an algorithm for finding the desired matching.

Since we know that stable matchings exist, it is an instructive and rather easy exercise to show that the set of stable matchings forms a lattice in the following sense: if $\mu$ and $\mu'$ are matchings, we write $\mu \underset{M}{\succsim} \mu'$ if every man likes his $\mu$-mate at least as well as his $\mu'$-mate. It turns out that this partial ordering of stable matchings is a lattice, as we will show shortly. Further, it is not hard to show that if $\mu \underset{M}{\succsim} \mu'$, then $\mu' \underset{W}{\succsim} \mu$, meaning that if all men prefer $\mu$ to $\mu'$, then all women prefer $\mu'$ to $\mu$ (these observations are due to J. H. Conway and are contained in [4]). As a corollary, since every finite lattice has a largest and a smallest element, it follows that among all stable matchings there is one which is preferred to all others by the men and another by the women. These will be called the *$M(W)$-optimal matchings*. The algorithms of [**2**] always arrive at one of these extreme matchings.

Some twenty years later, Dubins and Freedman considered the following question: suppose it is known that the matchmaker will impose the $M$-optimal stable matching and suppose some participants have full knowledge of the preference of all of the others. Will it ever be possible for an individual or group of individuals to obtain a preferred mate by falsifying preference lists? It was shown in [**1**] that there are cases in which a woman can do better by falsifying. On the other hand, the main result showed that no man or coalition of men can ever be made better off by falsifying their true preferences. The conclusion is that Machiavellian behavior on the part of the men is not profitable and they can do no better than to list their true preferences, regardless of what the women do. The purpose of the present paper is to give a detailed analysis of the strategic possibilities for the women. It will turn out that it will almost always pay for some of the women to be "Machiavellian", and under certain plausible assumptions we will describe their best competitive behavior.

**2. Some History.** The scenario described above involving marriages is, of course, rather fanciful, but there are genuine real world situations in which the matching problem comes up regularly. The problem of assigning students to colleges (which was the original motivation for these investigations) is an example. The difficulty here is that, as is so often the case, there are complicating factors such as the possibility of financial aid of various kinds. However, there is at least one situation where the methods of [**2**] are not only applicable but are actually being applied and have been for more than thirty years. The "National Residents Matching Program", headquarters in Evanston, Illinois, has the task each year of assigning graduates of medical schools throughout the country to hospital programs in which they are to fulfill a residency requirement. It turns out that the procedure used by NRMP is precisely the one described in [**2**], except that their procedure leads to the hospital-optimal rather than the student-optimal matching of [**2**]. The Dubins-Freedman Theorem shows that when the student optimal procedure is used the students can do no better than to list the hospitals they are interested in in their true order of preference. Under the NRMP system, on the other hand, the results presented here show that in general there will be some students who can get into a preferred hospital by suitably falsifying their preferences.

It should be pointed out that the student-college problem is more general than the marriage problem in that the colleges are "polygamous" and may admit any number of students up to some fixed quota $q$. Many of the properties of the marriage problem carry over to the polygamous case, including those to be presented here. However, not all results carry over to the more general case.

In particular, Roth has recently shown [**6**] by an example that even when the college optimal matching is used, as in the case of NRMP, it may still be possible for a college to get a better class by appropriate misrepresentation of preferences.

We now return to our main subject.

**3. Preliminaries.** In this section we present two propositions which will be needed later and which are of some interest in themselves. These are special cases of some results of [**3**]. The first is rather surprising. Recall that in general there may be many stable matchings for a given set of men and women. Our result asserts, however, that the people who are unmatched or, as we shall say, self-matched, are the same for all of these matchings. This is particularly striking in the context of the student-hospital situation. Suppose NRMP were to change its policy and impose the student rather than hospital-optimal matching. Of course this would make all the students at least as well off, but those students who were not accepted by any hospital would still not be accepted and, on the other hand, each hospital would end up admitting the same number of students, though in general not the same set, under the student-optimal as under the hospital-optimal scheme. This rather non-obvious fact turns out to be quite easy to prove. Our second proposition, by contrast, seems extremely plausible. It asserts that if additional women are added to the "pool", this can never make any of the men worse off. The proof of this, however, is not so straightforward.

We will need a little terminology. A *preference pattern* will consist of a triple $(M, W; P)$ where $M$ and $W$ are the men and women and $P$ represents their preferences, that is, the ordered lists of the members of $M \cup W$. Because of our convention that an unmatched person is considered to be self-matched, we may define a matching $\mu$ as a bijection of $M \cup W$ onto itself which is of order 2, that is, $\mu \cdot \mu =$ identity, and such that if $m$ or $w$ is not self-matched, then $\mu(m) \in W$ and $\mu(w) \in M$. We call $\mu(m)$ $(\mu(w))$ the *mate* of $m$ $(w)$ under $\mu$. We say that a pair $(m, w)$ *blocks* the matching $\mu$ if $m$ prefers $w$ to $\mu(m)$ and $w$ prefers $m$ to $\mu(w)$. The matching $\mu$ is then *stable* if it is not blocked by any pair. If $\mu$ and $\mu'$ are matchings, we say that $m$ $(w)$ prefers $\mu'$ to $\mu$ if he (she) prefers $\mu'(m)$ to $\mu(m)$ $(\mu'(w)$ to $\mu(w))$. We write $\mu \succsim_M \mu'$ if every man is at least as well off under $\mu$ as under $\mu'$. A key result is now the following.

LEMMA 1. *Suppose $W \subset W'$ and $\mu$ is a stable matching for $(M, W; P)$ and $\mu'$ for $(M, W'; P')$ where $P'$ agrees with $P$ on $W$. Let $M_\mu$ be all men who prefer $\mu$ to $\mu'$ and let $W_{\mu'}$ be all women who prefer $\mu'$ to $\mu$. Then $\mu$ and $\mu'$ are bijections between $M_\mu$ and $W_{\mu'}$.*

This is a sort of decomposition lemma represented schematically by the diagram below.

$$M_\mu \underset{\mu'}{\overset{\mu}{\substack{\leftrightarrow \\ \leftrightarrow}}} W_{\mu'}$$

$$M - M_\mu \underset{\mu'}{\overset{\mu}{\substack{\leftrightarrow \\ \leftrightarrow}}} W' - W_{\mu'}$$

It will suffice to show that $\mu(M_\mu) \subset W_{\mu'}$ and $\mu'(W_{\mu'}) \subset M_\mu$, for since $M_\mu$ and $W_{\mu'}$ are finite and $\mu$ and $\mu'$ are injective, they must be surjective as well. For $m$ in $M_\mu$ we know $\mu(m) \neq m$ since $m$ prefers $\mu(m)$ to $\mu'(m)$; so let $w = \mu(m)$. Then $w \in W$ and $w \neq \mu'(m)$. Further, $w$ prefers $\mu'(w)$ to $m$ for if not $(m, w)$ would block $\mu'$, so $w \in W_{\mu'}$. A symmetric argument shows that $\mu'(W_{\mu'}) \subset M_\mu$. □

Our first result is an immediate consequence of the lemma.

PROPOSITION 1. *If $\mu$ and $\mu'$ are stable matchings for $(M, W)$, then the people who are self-matched are the same for both.*

□ Suppose $\mu'(m) = m$ but $\mu(m) = w$. Then $m \in M_\mu$ but then, by the Lemma for the case $W = W'$, $m \in \mu'(W_{\mu'})$, contradicting $\mu'(m) = m$. □

LEMMA 2. *Let $\mu$ and $\mu'$ be as in Lemma 1 and define $\mu''$ to agree with $\mu$ on $M_\mu \cup W_{\mu'}$ and with $\mu'$ on $(M - M_\mu) \cup (W' - W_{\mu'})$. Then $\mu''$ is stable for $(M, W')$.*

□ Clearly $\mu''$ restricted to $M_\mu \cup W_{\mu'}$ or $(M - M_\mu) \cup (W' - W_{\mu'})$ is stable, so if there is a blocking pair, we must have either $m \in M_\mu$, $w \notin W_{\mu'}$ or $m \notin M_\mu$ and $w \in W_{\mu'}$. Suppose $w$ prefers $m$ to $\mu''(w)$.

*Case* 1. $m \in M_\mu$, $w \in W'_\mu - W_{\mu'}$. Then $m$ prefers $\mu(m)$ to $\mu'(m)$ whom he likes at least as well as $w$, so $(m, w)$ does not block $\mu''$.

*Case* 2. $m \notin M_\mu$, $w \in W_{\mu'}$. Then $m$ likes $\mu'(m)$ at least as well as $\mu(m)$ whom he prefers to $w$, since $\mu$ is stable so, again, $(m, w)$ does not block $\mu''$. □

The lattice property is a consequence of Lemma 2, for note that $\mu''$ assigns each man his favorite of $\mu(m)$ and $\mu'(m)$, so in usual lattice notation we write $\mu'' = \mu \vee \mu'$. As stated in the introduction, it now follows that for any $(M, W; P)$ there exist matchings $\mu_M$ and $\mu_W$ which are $M$-optimal and $W$-optimal, respectively. We now have

PROPOSITION 2. *With the hypothesis of Lemma 1 let $\mu_M$ be the M-optimal matching for $(M, W; P)$ and let $\mu'_M$ be the M-optimal matching for $(M, W'; P')$. Then every man is at least as well off under $\mu'_M$ as under $\mu_M$.*

□ From Lemma 2, $\mu_M \vee \mu'_M$ is stable for $(M, W'; P')$ and

$$\mu \underset{M}{\gtrsim} \mu' \vee \mu' \underset{W}{\gtrsim} \mu \quad . \square$$

**4. The Matching Game.** Since the Dubins-Freedman Theorem shows that it is never advantageous for the men to falsify their preferences, we will assume from here on that the men always submit their true preference list. For the special case where there is only one stable matching, the Dubins-Freedman Theorem implies that honesty is also the best policy for the women. In all other cases, however, this is not so.

THEOREM 1. *If there is more than one stable matching, then there is at least one woman who will be better off by falsifying, assuming the others tell the truth.*

□ By hypothesis $\mu_M \neq \mu_W$, so let $w$ be any woman such that $\mu_W(w) \neq \mu_M(w)$. Now let $w$ falsify by removing from her preference list all men who rank below $\mu_W(w)$. Clearly the matching $\mu_W$ will still be stable under these preferences (there are now fewer possible blocking pairs). Letting $\mu'_M$ be the $M$-optimal matching for these new preferences, it follows from Proposition 1 that $w$ is not self-matched by $\mu'_M$ and hence she is matched with someone she likes at least as well as $\mu_W(w)$, since all other men have been removed from her list, and she prefers $\mu_W(w)$ to $\mu_M(w)$ so she prefers $\mu'_M$ to $\mu_M$. □

Theorem 1 shows that the policy of honest revelation of preference is "unstable" for the women in a different sense of instability, namely, if each woman expects the others to be honest, it will in general be possible for at least one woman to improve her position by lying. This leads one to ask the following question: if honesty is not the best policy, is there any set of policies or "strategies" which have the property that once they are adopted by the women there will be no advantage to any one woman in changing her strategy? It will be useful here to introduce some standard game theoretic-terminology and we will henceforth refer to our model as the *matching game*. We are dealing here with perhaps the most important concept in game theory, that of an *equilibrium point*. An abstract game consists of a set of players each of whom has a certain set of *strategies*. In a play of the game each player selects one of her strategies and the "rules of the

game" then assign her some *payoff* which may be a number ("score") or, as in our case, a mate. A set of strategies, one for each player, forms an *equilibrium point* if no player can, by changing her strategy, achieve a better payoff, assuming the other players do not change theirs. The strategies form a *strong equilibrium point* if no subset of players by changing its strategies can achieve a better payoff for all of its members. Are there any equilibrium point strategies for the women in the matching game? The next two theorems answer this question.

THEOREM 2. *Let $\mu$ be any stable matching for $(M, W, P)$ and suppose each woman in $\mu(M)$ chooses the strategy of listing only $\mu(m)$ on her preference list. This is an equilibrium point.*

□ It is clear that $\mu$ is stable under these falsified preferences which we will denote by $P'$. Further, $\mu$ is the only stable matching for $(M, W; P')$, for any other matching would leave some $w$ in $\mu(M)$ unmatched, which is not possible by Proposition 1. Hence $\mu$ is the $M$-optimal matching for $(M, W; P')$.

To see that $P'$ is an equilibrium point, suppose some $w$ now changes her preference list leading to a new $M$-optimal matching $\mu'$ which gives her a mate $m' = \mu'(w)$ whom she prefers to $\mu(w)$ under true preference. Then $m'$ must have been matched by $\mu$ to some $w'$, for if not $(m', w)$ would have blocked $\mu$ in $(M, W; P)$. But then $w'$ is self-matched under $\mu'$ since $m'$ was the only man on her $P'$-list. This means $m'$ prefers $w$ to $w'$, but if this were so, again $(m', w)$ would have blocked $\mu$, a contradiction. □

Theorem 2 says that the women can force any matching $\mu$ which is stable under the true preferences by equilibrium point strategies. We now present a sort of converse.

THEOREM 3. *Suppose the women choose any set of strategies $P'_W$ (preference lists) that form an equilibrium point for the matching game. Then the corresponding $M$-optimal matching for $(M, W; P')$ is one of the stable matchings of $(M, W; P)$.*

This theorem has been proved by Roth [**5**] making use of the properties of the matching algorithm of [**2**]. We present here a direct proof.

□ Suppose $\mu'$ is the $M$-optimal matching for $(M, W; P')$ but $(m, w)$ blocks $\mu'$ under $w$'s true preference. We will show that $P'$ is not an equilibrium point. Namely, let $w$ refalsify by listing only $m$ on her preference list. Then she will get him, for let $\mu''$ be the $M$-optimal matching for the new preference $P''$. If $w$ does not get $m$, then she is self-matched by $\mu''$ so by stability of $\mu''$, $m$ prefers $\mu''(m)$ to $w$ and by assumption $m$ prefers $w$ to $\mu'(m)$, so $m$ prefers $\mu''(m)$ to $\mu'(m)$. But clearly the matching $\mu''$ would be stable for $(M, W - w; P')$, where $P'$ is restricted to $W - w$. Thus $m'$ is worse off under the $M$-optimal matching for $(M, W; P')$ than he is under $\mu''$ for $(M, W - w, P')$, which directly contradicts Proposition 2 of the previous section. □

To summarize this section, we see that by falsifying appropriately the women can achieve by equilibrium point strategies any stable matching, thus, in particular, the $W$-optimal matching. On the other hand, the women cannot get too greedy for if any set of strategies gives some woman $w$ a mate whom she likes better than $\mu_W(w)$, this will not be an equilibrium point, by Theorem 3, so some other woman can change the matching to her advantage by choosing a different strategy.

**5. Strong Equilibrium Points.** By Theorem 2, the women can achieve any stable matching $\mu$ by equilibrium point strategies. However, unless $\mu = \mu_W$ the equilibrium point will not be strong. To see this note that if $\mu \neq \mu_W$, then $\mu(w) \neq \mu_W(w)$ for at least two women, for say, $\mu(w_1) \neq \mu_W(w_1) = m_1$. By Proposition 1, $m_1$ is not self-matched by $\mu$ and hence $\mu(m_1) = \breve{w}_2$ and $\mu(w_2) = m_1 \neq \mu_W(w_2)$. To show $\mu$ is not the matching of a strong equilibrium point we let $\breve{W}$ be all $w$ such that $\mu(w) \neq \mu_W(w)$. Let all $w$'s in $\breve{W}$ refalsify by pretending $\mu_W(w)$ is the only man on their list. Then $\mu_W$ is stable for this new preference and hence since all $\breve{W}$ are matched by $\mu_W$, they are matched by the $M$-optimal matching for the new preference.

Do there exist strong equilibrium points? Yes.

THEOREM 4. *Let each woman $w$ submit a preference list in true order of preference but removing all men who are ranked below $\mu_W(w)$. These preferences $P'$ are a strong equilibrium point.*

□ We claim that $\mu_W$ is the only stable matching for $(M, W; P')$, for clearly $\mu_W$ is stable and any other stable matching $\mu'$ must have $\mu'(w) \neq \mu_W(w)$ for some $w$; hence if $w$ is not self-matched, then $\mu'(w)$ is preferred by $w$ to $\mu_W(w)$. Since $\mu_W$ is the $W$-optimal stable matching, this means $\mu'$ is unstable under true preferences, hence is blocked by some pair $(m, w)$. But by construction of $P'$, this means $(m, w)$ blocks $\mu'$ under $P'$ preferences, contradicting $P'$-stability of $\mu'$.

Now since $\mu_W$ is the only stable matching for $(M, W; P')$ it is the $W$-optimal matching for $P'$. If some subset $\widehat{W}$ could get a better payoff by falsifying, then it would get a better payoff than from the $W$-optimal matching of $(M, W; P')$, but by the Dubins-Freedman Theorem this is impossible. □

It seems reasonable to consider the falsification of $P'$ of Theorem 4 as the best method of play for the women. It is (a) a strong equilibrium point so no woman or set of women would be tempted to deviate from $P'$, and, (b) among all equilibrium point strategies it gives the women the highest possible payoff. It would have been nice to be able to assert that the matching $\mu_W$ is the only matching obtainable from a strong equilibrium point. Unfortunately this is not the case, as shown by the following example. The true preferences are given by Table 1.

TABLE 1

| | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
|---|---|---|---|---|
| $m_1$ | 2,2 | 1,3 | 3,$x$ | $x$,$x$ |
| $m_2$ | 1,3 | 3,2 | 2,3 | 4,$x$ |
| $m_3$ | $x$,$x$ | $x$,$x$ | 1.2 | 2,$x$ |
| $m_4$ | 3,1 | 4,1 | 2,1 | 1,1 |

TABLE 2

| | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
|---|---|---|---|---|
| $m_1$ | $x$ | 2 | $x$ | $x$ |
| $m_2$ | 2 | $x$ | 2 | $x$ |
| $m_3$ | $x$ | $x$ | 3 | $x$ |
| $m_4$ | 1 | 1 | 1 | 1 |

The first entry in box $i, j$ is the ranking of $w_j$ on the list of $m_i$. If it is $x$, this means that $w_j$ is not in the list of $m_i$. Thus, $m_2$ ranks $w_2$ in third place and $w_1$ is not in the list of $m_3$. The second entry is the ranking of $m_i$ by $w_j$. If it is $x$, this signifies that $m_i$ is not in the list of $w_j$. So, $w_1$ ranks $m_1$ in second place and $m_3$ is not in the list of $w_4$. For these preferences the matching $\mu_W$ is given by $(m_1, w_1)$, $(m_2, w_2)$, $(m_3, w_3)$, $(m_4, w_4)$. Now suppose all women use the system of preferences, $P'$, given by Table 2 above. The entry in box $i, j$ is the ranking of $m_i$ by $w_j$. The $M$-optimal matching for these new preferences is given by $(m_2, w_1)$, $(m_1, w_2)$, $(m_3, w_3)$, $(m_4, w_4)$. We assert that $P'$ is a strong equilibrium point. In fact, no subset $W' \subseteq W$ which contains $w_4$, by falsifying, can improve the situation of all its members, since the mate of $w_4$ is the best possible; if $W'$ contains $w_3$, it cannot improve the situation of $w_3$, for if so $(m_4, w_4)$ would block the new matching. If $W'$ contains $w_2$, it cannot get a better mate for $w_2$ by falsifying, for if so $(m_4, w_4)$ or $(m_2, w_3)$ would block the new matching. If $W'$ contains $w_1$, it is not possible to improve the payoff of $w_1$, for if so $(m_4, w_4)$ or $(m_1, w_2)$ would block the new matching.

**6. Dominated Strategies.** There is one more important game-theoretic concept which is illustrated by the Matching Game. Let $\sigma$ and $\sigma'$ be two strategies for some player $w$ in a general $n$-person game. We say that $\sigma$ *dominates* $\sigma'$ if the payoff to $w$ when she plays $\sigma$ is at least as high as when she plays $\sigma'$ no matter what strategies the other players play, and $\sigma$ *strictly dominates* $\sigma'$ if the payoff to $w$ when she plays $\sigma$ is higher than when she plays $\sigma'$ for at least one set of strategies for the other players. A strategy is called *dominant* if it dominates all other strategies.

As an illustration of these concepts, the Dubins-Freedman Theorem shows that for each man revealing true preferences is a dominant strategy. For the women there are no dominant strategies except for the special case where $|W| = 1$. If there is only one woman $w$, then revealing true preferences strictly dominates any other strategy, for if she submits her true preferences, she will get the highest man on her list who has listed her. If she falsifies by, for example, listing $m' \succ m$ instead of $m \succ m'$, she will be worse off in the case where $m$ and $m'$ are the only men who are willing to marry her. From now on it will be assumed that there are at least two women in $W$. We will show that the strategy of listing only one man, as in Theorem 2, is dominated unless that man happens to be the woman's true first choice. In fact, we can essentially characterize the dominated strategies of the women by the following theorems, where $a \underset{x}{\succ} b$ means that $x$ prefers $a$ to $b$ under the true preference $P_x$.

THEOREM 5. *Any strategy $P_w'$ in which $w$ does not list her true first choice at the head of her list is strictly dominated.*

□ This has been proved by Roth **[5]** using the matching algorithm. We present here a direct proof.

Let $P_w'$ be a strategy described above. We will show that $P_w'$ is strictly dominated by $P_w''$ which lists $m_1$ ($w$'s favorite man) in first place and leaves the rest of the list unchanged. Let $\mu'$ and $\mu''$ be the corresponding $M$-optimal matchings (the strategies of all other players are assumed unchanged). If $\mu''(w) = m_1$, there is nothing to prove so we suppose that $\mu''(w) \neq m_1$. Then

(1) $\mu''(m_1)$ is preferred by $m_1$ to $w$ or $w$ is not on $m_1$'s list,

for if not $(m_1, w)$ would block $\mu''$. Hence $\mu''$ is stable under preference $P_w''$. So,

$$\mu'(m_1) \underset{m_1}{\succ} \mu''(m_1), \tag{2}$$

from the $M$-optimality of $\mu'$. Furthermore, from (1) and (2) it follows that $\mu'(m_1) \underset{m_1}{\succ} w$, and since all the other elements different from $m_1$ are ranked in the same ordering in both lists $P_w'$ and $P_w''$, it follows that $\mu'$ is stable under $P_w''$, and hence $\mu''(w) = \mu'(w)$; so $w$ is no worse off using $P_w''$ than using $P_w'$.

To show that $P_w''$ strictly dominates $P_w'$, let $m'$ be the first element in $P_w'$. Consider the following preference pattern: $m'$: $\{w\}$, $m_1 = \{w\}$ and no other man lists $w$. Then we can see that $m_1 = \mu''(w) \underset{w}{\succ} \mu'(w) = m'$, which concludes the proof. □

Our final result states that Theorem 5 describes essentially all the dominated strategies.

THEOREM 6. *Let $P_w'$ be any strategy for $w$ in which* (a) *$m_1$ ($w$'s favorite man) is listed first,* (b) *$P_w'$ contains only men $m$ who are on $w$'s true preference list $P_w$. Then $P_w'$ is not dominated.*

□ We will show that for any other strategy $P_w''$ there exists strategies $\bar{P}$ for the other players such that $\mu'(w) \underset{w}{\succ} \mu''(w)$, where $\mu'$ and $\mu''$ are the $M$-optimal matchings for $(M, W; P_w' \cup \bar{P})$ and $(M, W, P_w'' \cup \bar{P})$, respectively. There are three cases. We first suppose $P_w''$ also satisfies (a) above.

*Case* 1. $P_w'$ contains some $m$ not on $P_w''$. Then for $\bar{P}$ we suppose that $m$ lists $w$ as first choice and no other man lists $w$. Then $\mu'(w) = m$ while $w$ is self-matched under $\mu''$ (here we used that $m \in P_w$).

*Case* 2. $P_w''$ contains some $m$ not on $P_w'$. Then for $\bar{P}$ suppose $m$ has preference list $\{w \succ w'\}$ for some $w'$, $m_1$ has preference list $\{w' \succ w\}$, and $w'$ has preference list $\{m \succ m_1\}$ and no other man lists $w$ or $w'$. Then one verifies that $\mu'(w) = m_1 \underset{w}{\succ} \mu''(w) = m$.

*Case* 3. Let lists $P_w'$ and $P_w''$ be the same but $w$ prefers $m$ to $m'$ in $P_w'$ and $m'$ to $m$ in $P_w''$. Then suppose preference list for $m_1$ is $\{w' \succ w\}$, for $m$ is $\{w \succ w'\}$, for $m'$ is $\{w \succ w'\}$, and for $w'$ is $\{m' \succ m_1 \succ m\}$, and no other men list either $w$ or $w'$. It is an instructive exercise to verify

that $\mu'(w) = m_1$ while $\mu''(w) = m'$.

We have seen that if $P'_w$ satisfies (a) and (b) and $P''_w$ satisfies (a), then for some $\breve{P}$, $\mu'(w) \underset{w}{\succ} \mu''(w)$. If $P''_w$ does not satisfy (a), then by Theorem 5 there is some $P'''_w$ which dominates $P''_w$ and $P'''_w$ satisfies (a) so we construct $\breve{P}$ so that $\mu'(w) \underset{w}{\succ} \mu'''(w)$ but $\mu'''(w) \underset{w}{\succsim} \mu''(w)$ by dominance. The proof is now complete. □

REMARK 1. Condition (b) above is needed to avoid cases in which $P'_w = \{m_1 \succ m\}$ where $m$ is not in $P_w$. It is clear that this is strictly dominated by $P''_w = \{m_1\}$.

REMARK 2. Note that the counterexample of Section 4 uses only undominated strategies for the women. Thus we do not have a unique strong equilibrium point for $W$ even when restricting to the use of undominated strategies.

### References

1. L. E. Dubins and D. Freedman, Machiavelli and the Gale-Shapley algorithm, this MONTHLY, 88 (1981) 485–494.
2. D. Gale and L. S. Shapley, College admissions and the stability of marriage, this MONTHLY, 69 (1962) 9–14.
3. D. Gale and M. Sotomayor, Some remarks on the stable matching problem, Operations Research Center report OCR-83-12, University of California, Berkeley, 1983.
4. D. E. Knuth, Mariages Stables, Montreal University Press, 1976.
5. A. E. Roth, Misrepresentation and stability in the marriage problem, Working paper #160, Department of Economics, University of Pittsburgh, Pennsylvania, 1983.
6. ______, Incentives in the college admissions problem and related two-sided markets, Working paper #173, Department of Economics, University of Pittsburgh, Pennsylvania, 1983.

---

## 144. MISCELLANEA

### Inflation

The growth of mathematics is producing discontent:
The inflation rate per annum is pushing 10 per cent.
Faced with so much information, it's not easy to succeed
In locating any theorem that you appear to need.
Our plethora of indices can leave you in the lurch,
For it takes less time to prove it than to go and make a search.
I can offer a solution, but it's totally upsetting:
We need to introduce a way of constantly forgetting
The results that won't be needed for another 20 years,
At the end of which they'll surface with appropriate loud cheers,
While the ones that won't be needed till forever and a day,
Once their authors get their tenure will be firmly thrown away.

—R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S(10-12).** A Tangle of Mathematical Yarns. Joe Kennedy, Diane Thomas. (P.O. Box 132, Oxford, OH 45056), 1983, $7.95 (P). An appealing collection of tall tales with embedded math problems, generally at the high-school level. Totally unrealistic, the problems use chance and corny wit to entice students into complex 3-5 paragraph situations requiring careful reading, sophisticated problem-solving skills, but only simple mathematics. LAS

**General, S(13), L**.** Mathematical Byways in Ayling, Beeling, and Ceiling. Hugh ApSimon. Oxford U Pr, 1984, xii + 97 pp, $9.95. [ISBN: 0-19-853201-6] A collection of intriguing problems involving very little formal mathematical knowledge (high school algebra suffices for most), but considerable ingenuity. Each is set in the following format: a specific problem is given followed by a general solution and particularized to the original problem. This is followed by a discussion about the difficulties in setting the problem, and finally the author gives an unresolved related problem (appropriate for undergraduate research perhaps, or personal recreation). LCL

**General, P.** Science, Computers, and the Information Onslaught: A Collection of Essays. Ed: Donald M. Kerr, et al. Academic Pr, 1984, xiii + 276 pp, $29.50. [ISBN: 0-12-404970-2] Proceedings of a June 1981 meeting at Los Alamos in which mathematicians, physicists, linguists and computer scientists met with leaders in science administration and national security to assess how developments in science and technology could help society cope with the vast quantities of data now being produced. LAS

**Elementary, T(13: 1).** Why Math?. R.D. Driver. Undergrad. Texts in Math. Springer-Verlag, 1984, xiv + 233 pp, $24. [ISBN: 0-387-90973-7] For a course in general mathematics. Emphasis on quantitative reasoning in arithmetic and algebra. Touches of binary arithmetic, sets, probability and cardinality. Two chapters on motion problems including some in special relativity. Excellent, but expensive, source of elementary real-world problems. Prerequisite is a "little algebra and geometry." JK

**Elementary, T(13), S.** Probability Without Tears. Derek Rowntree. Charles Scribner's, 1984, vi + 169 pp, $13.95. [ISBN: 0-684-17994-6] A self-paced "programmed" format, written for students with little mathematics background, covering elementary counting principles (the product role, the sum role), permutations and combinations. LCL

**Precalculus, T(13-14: 1).** Intermediate Algebra with Applications. Terry H. Wesner, Harry L. Nustad. Wm C Brown, 1985, xiii + 799 pp. [ISBN: 0-697-00117-2] Very readable text. Excellent book for students with some algebra background. Contains no trigonometry. Has many examples and illustrations. TR

**Precalculus, T(13-14: 1).** College Algebra and Trigonometry. Hal G. Moore, Adil Yaqub. Kendall/Hunt, 1984, x + 822 pp, $24.95 (P). [ISBN: 0-8403-3229-7] An informal text--easy to read. Contains many diagrams and examples. Emphasizes understanding of the underlying mathematics. De-emphasizes the use of tables in trigonometry and logarithms and encourages the use of calculators for any arithmetic calculations. TR

**History, S(16-17), P*, L*.** The Mathematics of Sonya Kovalevskaya. Roger Cooke. Springer-Verlag, 1984, xiii + 234 pp, $29.80. [ISBN: 0-387-96030-9] A fascinating mathematical biography, tracing in modest detail the evolution of Kovalevskaya's professional life as the nineteenth century's leading woman mathematician. Contains extensive quotations from correspondence (especially with her mentors Weierstrass and Mittag-Leffler), and detailed contextual analysis of Kovalevskaya's major mathematics papers. Several appendices bring the book within range of good undergraduates. LAS

**History, P, L**.** George Pólya: Collected Papers. Ed: Joseph Hersch, Gian-Carlo Rota. Mathematicians of our Time. MIT Pr, 1984. V. III: Analysis, x + 536 pp, $50 [ISBN: 0-262-16096-X]; V. IV: Probability: Combinatorics; Teaching and Learning in Mathematics, ix + 642 pp, $60. [ISBN: 0-262-

16097-8] Final two volumes of Pólya's collected papers: 58 analysis papers spanning Pólya's entire career; 31 papers on probability and combinatorics; and 18 papers on teaching and learning. Each volume concludes with brief commentaries on the papers and a complete Pólya bibliography keyed to the four volumes. (Volume I and Volume II, edited by R.P. Boas, were published in 1974; TR, April 1975.) LAS

**History, P, L*.** Essays in the History of Mathematics. Ed: Arthur Schlissel. AMS Memoirs, No. 298. AMS, 1984, v + 73 pp, $9 (P). Six papers from a special session at the January 1981 annual AMS meeting, offering the perspective of mathematics researchers on the origins of their fields of specialty: Dantzig on linear programming, Neyman and Price on stochastic theory of epidemics, Hoppensteadt on population biology, and more. LAS

**History, L*.** Elements of Algebra. Leonhard Euler. Transl: John Hewlett. Springer-Verlag, 1984, lx + 593 pp, $28. [ISBN: 0-387-96014-7] Reprint of the 1840 English translation of Euler's 1770 Treatise of Algebra, originally published in German by the Royal Academy of Sciences at Petersburg, together with an "Addition" by LaGrange on continued fractions and other methods. Written shortly after Euler became blind and dictated to a servant not tutored in mathematics, the work is described by Frances Horner in the preface as "next to Euclid's Elements the most perfect model of elementary writing of which the scientific world is in possession." Includes an introductory essay on Euler written by C. Truesdall and reprinted from his 1984 volume An Idiot's Fugitive Essays on Science. LAS

**Foundations, T?(17: 1), P.** Automated Theorem Proving. Wolfgang Bibel. Friedr. Vieweg & Sohn (US Distr: Heyden & Son), 1982, xiii + 293 pp, $27.50 (P). [ISBN: 3-528-08520-7] Readable introduction to automated deduction (proof procedures) and computational aspects of first order logic. A "connection" method is developed to give unified treatment to reasoning methods (resolution, natural deduction, rewriting). Some exercises. RM

**Graph Theory, S(17), P.** Graph Theory and Combinatorics. Ed: Béla Bollobás. Academic Pr, 1984, xvii + 328 pp, $50. [ISBN: 0-12-111760-X] Proceedings of a conference which was held at Trinity College, Cambridge, from March 21-25, 1983 to honor Paul Erdős on his seventieth birthday. Includes 29 research papers in graph theory and combinatorics. CEC

**Discrete Mathematics, T*(14-16), S*, L.** Applied Combinatorics. Alan Tucker. Wiley, 1984, xi + 447 pp, $31.95. [ISBN: 0-471-86371-8] The Second Edition of this very versatile text (First Edition, TR, October 1980) puts the graph theory chapters first and gives more attention to applications to computer science. New sections on recursive programs, formal languages and grammars, finite state machines, logical propositions and computational complexity. Additional programming exercises throughout the text. LCL

**Number Theory, T**(17), S, P, L.** p-adic Numbers, p-adic Analysis, and Zeta-Functions, Second Edition. Neal Koblitz. Grad. Texts in Math., No. 58. Springer-Verlag, 1984, xii + 150 pp, $24.80. [ISBN: 0-387-96017-1] Changes from the First Edition (TR, August-September 1978) include material on the Iwasawa logarithm and the p-adic gamma-function added to Chapter IV. A very helpful appendix of answers and hints to the exercises from the book. BH

**Number Theory, P.** Lecture Notes in Mathematics-1052: Number Theory. Ed: D.V. Chudnovsky, et al. Springer-Verlag, 1984, 309 pp, $15.50 (P). [ISBN: 0-387-12909-X] Expanded texts of lectures given at the New York Number Theory Seminar held at the City University of New York during the spring of 1982. BH

**Number Theory, T*(17: 1), S*.** Introduction to Elliptic Curves and Modular Forms. Neal Koblitz. Grad. Texts in Math., No. 97. Springer-Verlag, 1984, viii + 248 pp, $35. [ISBN: 0-387-96029-5] Chapter 1 carefully introduces the neophytes to doubly periodic functions and arithmetic on elliptic curves, related by Weierstrass' "pe" function. Chapter 2 studies the classical functions of Riemann (zeta), Hasse-Weil (L), etc. Chapter 3 investigates the modular forms of integer weight, and Chapter 4 those of half integer weight. Motivated throughout by the congruent number problem: roughly, "which integers are the area of some right triangle with rational sides?" Ends with one version of Tunnell's theorem. Numerous exercises, many with answers and copious hints. Prerequisites: first courses in complex analysis and Galois theory; some familiarity with projective spaces and Fourier transforms. YN

**Linear Algebra, S(14).** Applications of Linear Algebra, Third Edition. Chris Rorres, Howard Anton. Wiley, 1984, ix + 364 pp, $16.95 (P). [ISBN: 0-471-86800-0] For use during or following a standard sophomore course in linear algebra; to the previous applications in business, economics, physics, computer science, geometry, ecology, sociology, demography, genetics and linear programming, this edition adds chapters on cubic spline interpolation, cryptography, computed tomography. Solution manual available. (First Edition, TR, November 1977; Second Edition, TR, August-September 1980.) JNC

**Linear Algebra, T(13).** The Mathematics of Matrices: A First Book of Matrix Theory and Linear Algebra. Philip J. Davis. Robert Krieger, 1984, xiii + 348 pp, $26.50. [ISBN: 0-89874-756-2] Another entry in the growing field of linear algebra texts, this one written with possible use at the high school level in mind. (1973 Xerox Second Edition, TR, May 1975.) AWR

**Algebra, S(18), P.** Methods in Ring Theory. Ed: F. van Oystaeyen. NATO ASI Ser. C, V. 129. D Reidel, 1983, viii + 576 pp, $74. [ISBN: 90-277-1743-5] A collection of papers on various topics in

ring theory growing out of the NATO Advanced Study Institute on Methods in Ring Theory held in Antwerp, Belgium in 1983. List of participants, index. JS

**Algebra, P.** Lecture Notes in Mathematics-1083: Automorphic Forms on GL(3,R). Daniel Bump. Springer-Verlag, 1984, xi + 184 pp, $9.50 (P). [ISBN: 0-387-13864-1] The author lays a foundation for the study of automorphic forms on GL(3,R). Main topics of investigation include the study of Wittaker functions, Eisenstein series on GL(3), Ramanujan sums on GL(3), L-series, Hecke operators. LCL

**Finite Mathematics, T(13-14: 1).** Finite Mathematics with Applications. Shirley O. Hockett, Martin Sternstein. Robert Krieger, 1984, vi + 393 pp, $28.50. [ISBN: 0-89874-674-4] A straightforward no-nonsense introduction to most of the standard topics of finite mathematics, including probability, linear algebra, linear programming, and mathematics of finance. Appendix on set theory, tables. JS

**Complex Analysis, S(18), P.** Quadratic Differentials. Kurt Strebel. Erg. der Math. und ihrer Grenzgebiete, 3. Folge, Band 5. Springer-Verlag, 1984, xii + 184 pp, $38. [ISBN: 0-387-13035-7] "A quadratic differential on a Riemannian surface is locally represented by a holomorphic function element which transforms like the square of a derivative under a conformal change of the parameter... . A large part of this book is about the trajectory structure of quadratic differentials." MU

**Complex Analysis, T(18), P.** Linear Problems and Convexity Techniques in Geometric Function Theory. D.J. Hallenbeck, T.H. MacGregor. Mono. & Stud. in Math., No. 22. Pitman, 1984, xvii + 182 pp, $45. [ISBN: 0-273-08637-5] Begins with properties of univalent functions, develops topics in function algebras, and proves results (some new) involving linearity or convexity techniques. AWR

**Differential Equations, P.** Lecture Notes in Mathematics-1076: Infinite-Dimensional Systems. Ed: F. Kappel, W. Schappacher. Springer-Verlag, 1984, vii + 278 pp, $13.50 (P). [ISBN: 0-387-13376-3] The proceedings of the "Conference on Operator Semigroups and Applications" held in Styria, Austria, June 5-11, 1983. Twenty-two papers on recent advances, including, for instance, research on integral- and integro-differential equations, age-dependent population dynamics, approximation methods. LCL

**Differential Equations, P.** Operational Calculus: A Theory of Hyperfunctions. K. Yosida. Appl. Math. Sci., V. 55. Springer-Verlag, 1984, x + 170 pp, $22 (P). [ISBN: 0-387-96047-3] Revised and enlarged translation of Japanese edition published by University of Tokyo Press (1982). Develops theory of operational calculus using hyperfunctions and applies it to solve one-dimensional wave equation, telegraph equation, and heat equation. Additions to Japanese version include a simple proof of the Titchmarsh convolution theorem using Liouville's Theorem and an algebraic solution of Laplace's differential equation: $(a_2t+b_2)y'' + (a_1t+b_1)y' + (a_0t+b_0)y = 0$. Rather dry presentation. BH

**Partial Differential Equations, S(18), P.** Lecture Notes in Mathematics-1074: Weighted Inequalities and Degenerate Elliptic Partial Differential Equations. Edward W. Stredulinsky. Springer-Verlag, 1984, 143 pp, $7.50 (P). [ISBN: 0-387-13370-4] The first two chapters derive preliminary inequalities, involving mainly measure theory. The third chapter establishes the main result: the Hölder continuity of weak solutions of quasilinear degenerate elliptic systems. The author has sprinkled the text with welcome explanations that lead the reader through the technicalities; at times, however, the mathematics appears simpler than the syntax. YN

**Partial Differential Equations, S(17-18), P.** Bilinear Transformation Method. Yoshimasa Matsuno. Math. in Sci. & Engin., V. 174. Academic Pr, 1984, viii + 223 pp, $50. [ISBN: 0-12-480480-2] The bilinear transformation method is a tool for the solution of a wide class of nonlinear evolution equations. The first part of this monograph is an introduction to the method. The second part is a study of the mathematical structure of the Benjamin-Ono and related equations. AO

**Partial Differential Equations, P.** Lecture Notes in Mathematics-1075: Asymptotic Analysis for Integrable Connections with Irregular Singular Points. Hideyuki Majima. Springer-Verlag, 1984, ix + 159 pp, $9.50 (P). [ISBN: 0-387-13375-5] Uses "strongly asymptotic expansions of functions of several variables to prove existence theorems of asymptotic solutions to integrable systems of partial differential equations. ... Formulates and solves the Riemann-Hilbert-Birkhoff problem, and provides analogues of Poincaré's lemma and deRham cohomology theorem for integrable connections with irregular singular points." BH

**Numerical Analysis, P.** Large Scale Scientific Computation. Ed: Seymour V. Parter. Academic Pr, 1984, ix + 326 pp, $26. [ISBN: 0-12-546080-5] The invited lectures from a conference held at the University of Wisconsin, Madison, on June 17-19, 1983. These papers review recent developments in the solution of large-scale problems in fluid dynamics. AO

**Numerical Analysis, S(17-18), P, L.** Methods and Applications of Error-Free Computation. R.T. Gregory, E.V. Krishnamurthy. Texts & Mono. in Comp. Sci. Springer-Verlag, 1984, xii + 194 pp, $29.80. [ISBN: 0-387-90967-2] Theory of residue and p-adic arithmetic, leading to a system of "Hensel codes" that permits exact calculation of standard problems of linear algebra--suitable for ill-conditioned problems or as alternatives to numerically unstable algorithms. LAS

**Functional Analysis, P.** Ten Lectures on Operator Algebras. William Arveson. CBMS Reg. Conf. Ser. in Math., No. 55. AMS, 1984, viii + 93 pp, $14 (P). [ISBN: 0-8218-0705-6] Lectures on nonselfad-

joint algebras of operators on Hilbert spaces. Topics include triangular and quasitriangular operator algebras, the Feynman-Kac formula, commutative subspace lattices, reflexive and hyperreflexive operator algebras, and the absorption principle. Very readable and well-motivated. BH

**Functional Analysis, S(18), P.** Methods for Solving Incorrectly Posed Problems. V.A. Morozov. Transl: Z. Nashed, A.B. Aries. Springer-Verlag, 1984, xviii + 257 pp, $36 (P). [ISBN: 0-387-96059-7] A fine translation of the 1974 Russian edition, emphasizing "regular methods" for solving a variety of problems about operators on Hilbert space, problems which do not quite satisfy the definition of well-posed problems. Includes a section on "pseudosolutions." MU

**Analysis on Manifolds, P.** Semimartingales and their Stochastic Calculus on Manifolds. Laurent Schwartz. Pr U Montreal, 1984, 187 pp, $17 (P). [ISBN: 2-7606-0660-0] An exposition based on lectures given by Laurent Schwartz at the Université de Montréal in 1982. LCL

**Analysis, P.** An Infinitesimal Approach to Stochastic Analysis. H. Jerome Keisler. Memoirs of AMS, No. 297. AMS, 1984, x + 184 pp, $16 (P). Uses nonstandard analysis to study stochastic integral equations with respect to a Brownian motion. Does not presuppose prior knowledge of nonstandard analysis or the theory of stochastic integral equations. AO

**Analysis, P.** A Computer-Assisted Proof of Universality for Area-Preserving Maps. J.-P. Eckmann, H. Koch, P. Wittwer. AMS Memoirs, No. 289. AMS, 1984, vi + 122 pp, $12 (P). An existence theorem concerning the Feigenbaum constant 4.66920... associated with one-parameter families such as $T_a(x) = 1 - ax^2$ is proved using a computer program to verify the norm of a high-degree polynomial. The program is carefully crafted to take into account round-off error, and the authors provide extensive analysis of other possible sources of error (hardware, software, operating system) enroute to a claim that this really is a mathematics proof. LAS

**Geometry, S(15-17), L.** Problems in Geometry. Marcel Berger, et al. Transl: Silvio Levy. Problem Books in Math. Springer-Verlag, 1984, viii + 266 pp, $29.80. [ISBN: 0-387-90971-0] Written in response to request of readers of Marcel Berger's wide-ranging and concise Geometry for hints and solutions to problems therein. In 1985, look for the Springer-promised English edition of the five-volumed, well-received Geometry. JK

**Topological Groups, T(18: 1), S, L.** Matrix Groups, Second Edition. Morton L. Curtis. Universitext. Springer-Verlag, 1984, xiii + 209 pp, $19.80 (P). [ISBN: 0-387-96074-0] Apart from minor editorial changes, the only difference from the First Edition (TR, May 1980) is the inclusion of a short chapter on roots of a compact connected Lie group. JS

**Topological Groups, P.** Lecture Notes in Mathematics-1077: Lie Group Representations III. Ed: R. Herb, et al. Springer-Verlag, 1984, xi + 454 pp, $22.50 (P). [ISBN: 0-387-13385-2] Last of three volumes of articles by the main speakers at the University of Maryland's 1982-83 Special Year on Lie Group Representations. These papers come from Periods III, IV and V of the special year. LAS

**Operations Research, P.** Minimal Cost Flow in Processing Networks, A Primal Approach. J. Koene. CWI Tract, No. 4. Math Centrum, 1983, iv + 157 pp, Dfl. 22.70 (P). [ISBN: 90-6196-270-6] Processing networks can be used to model certain refining and blending processes. This monograph presents special techniques for solving minimum cost flow problems in these special network models. AO

**Operations Research, P.** Algorithms and Approximations for Queueing Systems. M.H. van Hoorn. CWI Tract, No. 8. Math Centrum, 1984, 122 pp, Dfl. 17.90 (P). [ISBN: 90-6196-274-9] A research tract that tries to bridge the gap between the theory of queueing and efforts to use that theory in numerical applications. AWR

**Game Theory, T(15), S, L.** Games, Theory and Applications. L.C. Thomas. Ser. in Math. & Its Applic. Halsted Pr, 1984, 279 pp, $65.95. [ISBN: 0-470-27507-3] An attempt to make the ideas of game theory accessible to a wide audience--undergraduate numerates--to use the author's designation. More mathematical than popularized descriptions (readers need some calculus, some visual acquaintance with matrices, basic probability), less demanding than a book including proofs or assuming background in linear programming or econometrics. Very nice balance, attractive text, with problems. AWR

**Optimization, T(17).** Optimization Theory and Applications. Jochen Werner. Adv. Lect. in Math. Heyden & Son, 1984, vii + 233 pp, $16 (P). [ISBN: 3-528-08594-0] An augmented version of lecture notes for an optimization course at Göttengen. Necessary tools from functional analysis are developed so as to enable emphasis on duality theory in convex programming and necessary optimality conditions for nonlinear optimization problems. No exercises. AWR

**Probability, T(16-18), S, P, L.** An Introduction to Probability Theory. P.A.P. Moran. Oxford U Pr, 1984, 542 pp, $19.95 (P). [ISBN: 0-19-853242-3] A moderately advanced introduction using measure-theoretic language and random processes, written to provide a background for more abstract and generalized treatises. Applications include discussions of Markov processes, Brownian motion, and random walks. LCL

**Probability, P.** Theory of Probability, Third Edition. Harold Jeffreys. Oxford U Pr, 1983, xi + 459 pp, $24.95 (P). [ISBN: 0-19-853193-1] Paperback reprint of the 1967 impression, which contained some corrections and amplifications, of the author's 1961 Third Edition. A classic treatise, originally published in 1939. RSK

**Probability, T(16-17: 2).** Mathematical Techniques of Applied Probability. Jeffrey J. Hunter. Operations Res. & Industrial Eng. Academic Pr, 1983. Volume 1: Discrete Time Models: Basic Theory, xiii + 239 pp, $32 [ISBN: 0-12-361801-0]; Volume 2: Discrete Time Models: Techniques and Applications, xiii + 286 pp, $35. [ISBN: 0-12-361802-9] Rigorous and thorough introduction, presuming a background in introductory probability, real analysis and linear algebra. Volume 1 presents the recurrent event model and introduces Markov chains, together with prerequisite mathematics. Volume 2 develops key properties of Markov chains and relates them to two main application areas, branching chains and discrete time queueing models. RSK

**Probability, T(16-17: 2).** Introduction aux Probabilités. P. Brémaud. Springer-Verlag, 1984, xv + 334 pp, $16 (P). [ISBN: 0-387-13612-6]

**Probability, P.** Stochastic Analysis and Applications. Ed: Mark A. Pinsky. Advances in Prob., V. 7. Dekker, 1984, x + 460 pp, $59.95. [ISBN: 0-8247-1906-9] Sixteen papers from leading authorities exhibit current research in, for example, stochastic integration, stochastic differential equations, stochastic optimization, stochastic problems in physics and biology. LCL

**Probability, T(13).** Introduction to Probability, Revised Edition. Nelson G. Markley. Ginn Custom, 1984, 300 pp, $13.95 (P). [ISBN: 0-536-04704-9] A nonthreatening introduction to elementary probability (including permutations, combinations, conditional probability, binomial distribution, normal distribution). No theorems or proofs, but lots of examples and related exercises. LCL

**Probability, P.** Lecture Notes in Economics and Mathematical Systems-231: The M/M/$\infty$ Service System with Ranked Servers in Heavy Traffic. G.F. Newell. Springer-Verlag, 1984, xi + 126 pp, $8.50 (P). [ISBN: 0-387-13377-1] Analysis of the distribution of the number of busy servers among the first s at time t for systems of infinitely many (preferentially ordered) servers in parallel. Closed form solutions, approximations and asymptotic analysis based on deterministic flow and diffusion models. RM

**Statistics, S(14-18), L.** Lecture Notes in Statistics-23: Estimation of Victimization Prevalence Using Data from the National Crime Survey. Diane Griffin Saphire. Springer-Verlag, 1984, v + 165 pp, $10.50 (P). [ISBN: 0-387-96020-1] Thorough analysis of different estimation models (both ad hoc and distribution-based) based on monthly household data from the national crime survey. A good case study in alternative means of statistical inference. LAS

**Statistics, S.** New Cambridge Elementary Statistical Tables. D.V. Lindley, W.F. Scott. Cambridge U Pr, 1984, 80 pp, $7.95 (P). [ISBN: 0-521-26922-9] Updated version of Cambridge Elementary Statistical Tables by Lindley and Miller, 1953. Contains tables for the most common parametric and nonparametric tests. RSK

**Statistics, P*.** Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices. Ludovic Lebart, Alain Morineau, Kenneth M. Warwick. Transl: Elisabeth Moraillon Berry. Wiley, 1984, xvi + 231 pp, $34.95. [ISBN: 0-471-86743-8] In the Wiley Series in Probability and Mathematical Statistics. Translation, with new examples and updated references, of the 1977 French edition entitled Techniques de la Description Statistique. Presents recently developed statistical techniques that are particularly suited to the description and analysis of large data sets. Presumes linear algebra and classical statistics. Good set of references. RSK

**Statistics, P.** Lecture Notes in Statistics-22: Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data. Søren Johansen. Springer-Verlag, 1984, viii + 126 pp, $12 (P). [ISBN: 0-387-90968-0] Theoretical investigation of linear and non-linear models motivated by data from a classic experiment involving the metabolism of ethanol in the liver. LAS

**Statistics, P.** Lectures in Computational Statistics. Ed: J.M. Chambers, et al. Compstat Lectures, No. 2. Physica-Verlag, 1984, 154 pp, $21 (P). [ISBN: 3-7051-0004-1] Algorithms for "parsimonious data fitting," the selection of mathematical models of exploratory data analysis with as few free parameters as possible. Designed for those who earn their living fitting data under constraints of resources and time. LAS

**Statistics, P.** An Introduction to the Theory of Large Deviations. D.W. Stroock. Universitext. Springer-Verlag, 1984, vii + 196 pp, $18 (P). [ISBN: 0-387-96021-X] Technical monograph based on a course given by the author. Main parts include extensions of Cramér's fundamental theorem and explanations of some recent results in the theory of large deviations from ergodic phenomena. RSK

**Statistics, T*(18: 2), P*.** Multivariate Observations. G.A.F. Seber. Wiley, 1984, xx + 686 pp, $49. [ISBN: 0-471-88104-X] In the Wiley Series in Probability and Mathematical Statistics. Presents both modern data-oriented techniques in multivariate analysis and classical methods supported by comments about robustness and general practical applicability. Presumes matrix algebra and familiarity with the multivariate normal distribution, multiple linear regression, and simple analysis of variance and covariance. Extensive references. RSK

**Statistics, P*.** Statistics: An Appraisal. Ed: H.A. and H.T. David. Iowa St U Pr, 1984, x + 664 pp, $31.25. [ISBN: 0-8138-1721-8] Proceedings of an international conference held in Ames, Iowa, in June 1983, to mark the 50th anniversary of the Statistical Laboratory at Iowa State University. Established by George Snedecor in 1933, it was the first such unit in the U.S. Papers deal with past achievements and future trends in major areas of statistical research. RSK

**Statistics.** Lecture Notes in Statistics-24: An Introduction to Bispectral Analysis and Bilinear Time Series Models. T. Subba Rao, M.M. Gabr. Springer-Verlag, 1984, viii + 280 pp, $15.50 (P). [ISBN: 0-387-96039-2]

**Statistics, P.** The Analysis of Categorical Data. Ed: A.K. Gupta. Comm. in Stat.: Theory & Methods, V. 12, No. 11. Dekker, 1983, 117 pp, (P). Communications in Statistics currently consists of five journals, Theory and Methods, Simulation and Computation, Sequential Analysis, Econometric Reviews and Statistical Reviews. Theory and Methods, issued 25 times a year, "focuses primarily on papers describing new applications of known statistical methods...and on articles dealing with a strong mathematical orientation that are significant to statistical studies." This special issue contains six research articles in the area of categorical data analysis. RSK

**Statistics, T(17: 1), P*.** Analysis of Ordinal Categorical Data. Alan Agresti. Wiley, 1984, ix + 287 pp, $35.95. [ISBN: 0-471-89055-3] In the Wiley Series in Probability and Mathematical Statistics. Contemporary treatment of methods of analyzing cross-classification tables having ordered, rather than just nominal, categories for at least one of the classifications. Includes an introduction to basic descriptive and inferential methods for categorical data. Good set of references. RSK

**Statistics, P, L*.** Applied Bayesian and Classical Inference: The Case of The Federalist Papers. Frederick Mosteller, David L. Wallace. Ser. in Statistics. Springer-Verlag, 1984, xxxviii + 303 pp, $29.80. [ISBN: 0-387-90991-5] Reprint of the original edition (published in 1964 by Addison-Wesley) supplemented by a 25-page Analytical Table of Contents containing descriptions of the contents of each section and a new concluding chapter reviewing authorship studies in the past 20 years. LAS

**Computer Literacy, S, L.** Overcoming Computer Illiteracy: A Friendly Introduction to Computers. Susan Curran, Ray Curnow. Penguin Books, 1983, 458 pp, $12.95 (P). [ISBN: 014-007159-8] An inviting, well-written, substantial introduction to computers, beginning with abstract data and real electronics: the emphasis is on hardware (not keyboards and printers but gates and busses), and machine-level code. Concludes with introductions to high-level languages, artificial intelligence, and expert systems. A significant, readable treatment intended for the curious layman. LAS

**Computer Literacy, T(13-18: 1), S.** An Introduction to Computers and Information Processing. Robert A. and Nancy Stern. Wiley, 1982, xxii + 637 pp, $9.80 (P). [ISBN: 0-471-09896-5] A business-oriented text. Touches on many basic topics in the area of computer literacy and computer "awareness." Also contains some good information about various languages, program logic, and system design. Includes an interesting chapter on the historical perspective of computers. TR

**Computer Literacy, T(13).** Computer Fundamentals for an Information Age. Gary B. Shelly, Thomas J. Cashman. Anaheim, 1984, xviii + 474 pp, $23.95 (P) [ISBN: 0-88236-125-2]; Workbook and Study Guide with Computer Software Projects, xiii + 321 pp, $10.95 [ISBN: 0-88236-126-0]; Transparency Masters, 206 pp. Glossy full-color presentation of computer vocabulary, covering hardware and business applications. BASIC is introduced in an appendix as an advanced feature; the word "algorithm" is not listed in the 17-page index. Extensive teaching and review aids: vocabulary lists, chapter summaries, workbook, review of questions, and transparency masters. Much ado about very little. LAS

**Computer Programming, S, P.** Introduction to PEARL: Process and Experiment Automation, Realtime Language: Description with Examples, Second Edition. Wulf Werum, Hans Windauer. Heyden & Son, 1983, xi + 183 pp, $20 (P). [ISBN: 3-528-13590-5] PEARL is a special-purpose programming language designed to meet the needs of real-time applications in the areas of process control and laboratory experimentation. PEARL was developed in Germany in the early 1970's and is now running on dozens of different computers. This text contains a description of the complete PEARL language syntax. MS

**Computer Science, P.** Computer Aided Analysis and Optimization of Mechanical System Dynamics. Ed: Edward J. Haug. NATO ASI Ser. F, V. 9. Springer-Verlag, 1984, xxii + 700 pp, $49.50. [ISBN: 0-387-12887-5] Proceedings of the NATO-NSF-ARO study institute held in Iowa City, Iowa, August 1-12, 1983. JAS

**Computer Science, T(13-14), S.** Common LISP, The Language. Guy L. Steele, Jr. Digital Pr, 1984, xii + 465 pp, $22 (P). [ISBN: 0-932376-41-X] LISP is a very popular programming language quite useful in the areas of artificial intelligence and natural language processing. One of the problems in using LISP, however, is that there are a number of incompatible dialects of the language. This text describes another dialect, called "Common LISP," which is intended to serve as a common dialect of the language across a wide range of different computer systems. This book is a syntactic and semantic reference manual for the language. MS

**Computer Science, P.** Models of the Lambda Calculus. C.P.J. Koymans. CWI Tract, No. 9. Math Centrum, 1984, iii + 181 pp, Dfl. 26.20 (P). [ISBN: 90-6196-275-7] Author's Ph.D. thesis on model theory of the pure untyped lambda calculus (the function space theory underlying algorithms). Follows categorical approach of Scott; discussion of hypergraph models. RM

**Computer Science, P.** Lecture Notes in Computer Science-178: Readings on Cognitive Ergonomics--Mind and Computers. Ed: G.C. van der Veer, et al. Springer-Verlag, 1984, 269 pp, $13 (P). [ISBN: 0-387-13394-1] Proceedings of the second European conference held in Gmunden, Austria, September 10-14, 1984. JAS

**Computer Science, P.** Lecture Notes in Medical Informatics-24: Medical Informatics Europe 84. Ed: F.H. Roger, et al. Springer-Verlag, 1984, xxvii + 778 pp, $34.50 (P). [ISBN: 0-387-13374-7] Proceedings of the congress held in Brussels, Belgium, September 10-13, 1984. JAS

**Computer Science, P.** Ninth Colloquium on Trees in Algebra and Programming. Ed: B. Courcelle. Cambridge U Pr, 1984, 326 pp, $39.50. [ISBN: 0-521-26750-1] Proceedings of a conference held in Bordeaux, France in March, 1983. The text contains reprints of 22 research papers on a wide range of theoretical computer science including data structures, regular expressions, lambda calculus, finite automata, and formal languages. The papers are all quite advanced and require a deep understanding of theoretical computer science and discrete mathematics. MS

**Computer Science, P.** Lecture Notes in Computer Science-179: How to Multiply Matrices Faster. Victor Pan. Springer-Verlag, 1984, xi + 212 pp, $11 (P). [ISBN: 0-387-13866-8] Surveys the progress made since 1978 in the development of fast algorithms for matrix multiplication and related problems from combinatorial analysis and linear algebra. AO

**Computer Science, S, P.** Lecture Notes in Computer Science-180: Ada Software Tools Interfaces. Ed: Peter J.L. Wallis. Springer-Verlag, 1984, 163 pp, $10 (P). [ISBN: 0-387-13878-1] Ada is a new programming language used primarily in the area of system's software. This text describes research work being conducted in building the tools (i.e., support programs) that will enhance the Ada programming environment and make the Ada programmer more productive. It contains 10 research papers which were presented at a Conference of Ada software tools in Bath, England in 1983. The reader should already be familiar with the syntax and semantics of the Ada language. MS

**Computer Science, P*, L.** The Carnegie-Mellon Curriculum for Undergraduate Computer Science. Ed: Mary Shaw. Springer-Verlag, 1985, 198 pp, $18.50 (P). [ISBN: 0-387-96099-6] Philosophy, objectives, requirements, and course descriptions for CMU's recently redesigned undergraduate curriculum in computer science. Relying on integration of topics to streamline the major, it requires only 10 computer science and 4 mathematics courses, reserving nearly 50% of the curriculum for liberal arts and elective non-CS courses. The volume contains reprints of several papers dealing with the curricular relationship between mathematics and computer science. LAS

**Control Theory, P.** Control and Dynamic Systems: Advances in Theory and Applications, V. 21: Nonlinear and Kalman Filtering Techniques, Part 3 of 3. Ed: C.T. Leondes. Academic Pr, 1984, xx + 422 pp, $49. [ISBN: 0-12-012721-0] A selection of new and substantive examples of the techniques and technology in the applications of nonlinear filters and Kalman filters. LCL

**Control Theory, T(18), S, P.** Differential Inclusions: Set-Valued Maps and Viability Theory. Jean-Pierre Aubin, Arrigo Cellina. Grund. der math. Wissenschaften, B. 264. Springer-Verlag, 1984, xiii + 342 pp, $44. [ISBN: 0-387-13105-1] A technical report on the motivations, ideas, and tools used in the study of dynamical systems in which a differential equation $x' = f(x)$ is replaced by a differential inclusion $x' \in F(x)$, where $F(x)$ is an associated set-valued map ("feasible velocities"). LCL

**Control Theory, P.** Impulse Control and Quasi-Variational Inequalities. Alain Bensoussan, Jacques-Louis Lions. Gauthier-Villars, 1984, xiv + 684 pp, $76. [ISBN: 2-04-015577-5] The general goal of the book is to explore analogies between variational inequalities (see the author's Applications of Variational Inequalities in Stochastic Control, TR, August-September 1979), and quasi-variational inequalities. LCL

**Systems Theory, P.** Simulation and Model-Based Methodologies: An Integrative View. Ed: Tuncer I. Ören, Bernard P. Zeigler, Maurice S. Elzas. NATO ASI Ser. F, V. 10. Springer-Verlag, 1984, xiii + 651 pp, $57.50. [ISBN: 0-387-12884-0] Invited addresses and contributed papers at a 1982 NATO Advanced Study Institute held at Ottawa, the goal of which was to foster simulation as a method applicable in many disciplines. AWR

**Applications (Biology), P, L.** Mathematical Models of Renewable Resources. Ed: Roland H. Lamberson. Humboldt State U (Math. Modeling Group), 1983, 245 pp, (P). Proceedings of a conference held at the University of Victoria in May 1983, primarily devoted to models of fisheries. LAS

**Applications (Biology), T(16).** Modeling Dynamic Phenomena in Molecular and Cellular Biology. Lee A. Segel. Cambridge U Pr, 1984, xx + 300 pp, $49.50; $12.95 (P). [ISBN: 0-521-25465-5; 0-521-27477-X] Intended for biology students who have studied calculus for one year; aims to show that analysis of mathematical models can give insight into biological problems. AWR

**Applications (Economics), P.** Coping with Complexity: Perspectives for Economics, Management and Social Sciences. Hans W. Gottinger. Theory & Decision Lib., V. 33. D Reidel, 1983, xv + 224 pp, $44. [ISBN: 90-277-1510-6] Most socio-economic systems are too complex to be modelled by a few equations but are not large enough to assume continuity. Gottinger proposes modelling these systems with sequential machines. Motivated by concept of complexity in automata theory, introduces design, control and evolution complexity to study stability and controllability of systems. KS

**Applications (Engineering), P.** Reliability Theory and Models: Stochastic Failure Models, Optimal Maintenance Policies, Life Testing, and Structures. Ed: Mohamed S. Abdel-Hameed, Erhan Çinlar, Joseph Quinn. Notes & Reports in Comp. Sci. & Appl. Math., V. 10. Academic Pr, 1984, xiii + 303 pp, $37.50. [ISBN: 0-12-041420-1] Proceedings of a conference held at Charlotte, North Carolina, June 24-26, 1983. Twenty papers on stochastic failure models, optimal maintenance and replacement policies, life testing, reliability structures, computability, and approximations. AO

**Applications (Engineering), P.** Flexible Shells: Theory and Applications. Ed: E.L. Axelrad, F.A. Emmerling. Springer-Verlag, 1984, ix + 282 pp, $23. [ISBN: 0-387-13526-X] Proceedings of the European Mechanics Colloquium 165. Contains seventeen papers on the theory of shells capable of large elastic displacements. AO

**Applications (Image Reconstruction), S(17), P.** Inverse Problems. Ed: D.W. McLaughlin. SIAM AMS Proc., V. 14. AMS, 1984, vii + 189 pp, $33. [ISBN: 0-8218-1334-X] This symposium (April 1983, New York) concerns the mathematics pertinent to: integral equations in geophysics; Radon transforms in radiation oncology; spectral techniques and boundary measurements in physics; probabilistic methods of maximal entropy to enhance blurred images. Interesting survey of the field and its open problems. YN

**Applications (Physics), S(18), P.** Differential Geometric Methods in Mathematical Physics. Ed: S. Sternberg. Math. Physics Stud. D Reidel, 1984, vii + 295 pp, $49. [ISBN: 90-277-1781-8] Research-expository papers delivered at a conference in Tel Aviv in August 1982, focusing more on differential geometry than on physics. A few concern pure symplectic geometry (e.g., Guillemin and Sternberg's), some have a more electromagnetic Yang-Mills flavor (Carmeli and Huleihil's), and a couple provide a colorful flow of explanations on quarks and plasmas for the non-physicists (Bleuler's and M.E. Mayer's). This variety of styles and topics will captivate both mathematicians and physicists. YN

**Applications (Physics), S(16-18), P*, L*.** Rational Thermodynamics, Second Edition. C. Truesdell. Springer-Verlag, 1984, xvii + 578 pp, $69. [ISBN: 0-387-90874-9] A significant enlargement of the 1969 McGraw-Hill First Edition featuring an extensive new historical introduction plus 22 appendices by individuals whose work has advanced the subject during the last fifteen years. "Rational thermodynamics is almost as much a restoration as it is a new science... . Taken as a whole [it] is neither more nor less axiomatic than any other mathematical science as treated by mathematicians." LAS

**Applications (Physics), P.** Chaos and Statistical Methods. Ed: Y. Kuramoto. Ser. in Synergetics, V. 24. Springer-Verlag, 1984, xi + 273 pp, $32. [ISBN: 0-387-13156-6] Proceedings of the Sixth Kyoto Summer Institute, devoted to chaos and statistical mechanics, held in Kyoto, Japan, in September 1983. Contains 36 papers presenting new results and serving as a review of the present status of this rapidly developing area. RSK

**Applications (Physics), T(17-18), S, P, L.** Group Theory in Physics. J.F. Cornwell. Techniques of Physics, V. 7. Academic Pr, 1984. Volume I, xvi + 371 pp, $75 [ISBN: 0-12-189801-6]; Volume II, xvi + 558 pp, $100. [ISBN: 0-12-189802-4] A self-contained introduction to that part of group theory that has been most useful in theoretical physics, using notation familiar to physicists, together with major applications to the study of molecular vibrations, atomic physics, and elementary particles. Note price! LCL

**Applications (Physics), T(17-18), P.** Diffraction for Materials Scientists. Jerold M. Schultz. Intern. Ser. in Physic. & Chem. Engin. Sci. Prentice-Hall, 1982, xii + 287 pp, $39.95. [ISBN: 0-13-211920-X] A textbook aimed at teaching the fundamentals of scattering theory and its universality and utility in materials science. The concepts of classical crystallography are deferred to the latter half of the book until after a theoretical framework is established. LCL

**Applications (Physics), S, L.** Energy: Facts and Figures. Robert H. Romer. Spring Street Pr (104 Spring St., Amherst, MA 01002), 1985, 68 pp, $8.95 (P). [ISBN: 0-931691-17-6] A fascinating collection of data and conversion factors for energy consumption, including data on energy, production and consumption for the past century. All tables and graphs; virtually no text. Tidbits: a walking person gets about 470 mpg; a jelly donut is worth $10^6$ joules. LAS

**Applications (Physics), T(18: 1, 2), P.** Gauge Theory of Elementary Particle Physics. Ta-Pei Cheng, Ling-Fong Li. Clarendon Pr, 1984, xi + 536 pp, $39. [ISBN: 0-19-851956-7] Modern theories of the strong, weak, and electromagnetic interactions are based on the principle of gauge symmetry. This work is an introduction to the principal ideas of gauge theories and their applications to elementary particle physics. AO

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

that $\mu'(w) = m_1$ while $\mu''(w) = m'$.

We have seen that if $P'_w$ satisfies (a) and (b) and $P''_w$ satisfies (a), then for some $\breve{P}$, $\mu'(w) \underset{w}{\succ} \mu''(w)$. If $P''_w$ does not satisfy (a), then by Theorem 5 there is some $P'''_w$ which dominates $P''_w$ and $P'''_w$ satisfies (a) so we construct $\breve{P}$ so that $\mu'(w) \underset{w}{\succ} \mu'''(w)$ but $\mu'''(w) \underset{w}{\succsim} \mu''(w)$ by dominance. The proof is now complete. □

REMARK 1. Condition (b) above is needed to avoid cases in which $P'_w = \{m_1 \succ m\}$ where $m$ is not in $P_w$. It is clear that this is strictly dominated by $P''_w = \{m_1\}$.

REMARK 2. Note that the counterexample of Section 4 uses only undominated strategies for the women. Thus we do not have a unique strong equilibrium point for $W$ even when restricting to the use of undominated strategies.

### References

1. L. E. Dubins and D. Freedman, Machiavelli and the Gale-Shapley algorithm, this MONTHLY, 88 (1981) 485–494.
2. D. Gale and L. S. Shapley, College admissions and the stability of marriage, this MONTHLY, 69 (1962) 9–14.
3. D. Gale and M. Sotomayor, Some remarks on the stable matching problem, Operations Research Center report OCR-83-12, University of California, Berkeley, 1983.
4. D. E. Knuth, Mariages Stables, Montreal University Press, 1976.
5. A. E. Roth, Misrepresentation and stability in the marriage problem, Working paper #160, Department of Economics, University of Pittsburgh, Pennsylvania, 1983.
6. ______, Incentives in the college admissions problem and related two-sided markets, Working paper #173, Department of Economics, University of Pittsburgh, Pennsylvania, 1983.

---

## 144. MISCELLANEA

### Inflation

The growth of mathematics is producing discontent:
The inflation rate per annum is pushing 10 per cent.
Faced with so much information, it's not easy to succeed
In locating any theorem that you appear to need.
Our plethora of indices can leave you in the lurch,
For it takes less time to prove it than to go and make a search.
I can offer a solution, but it's totally upsetting:
We need to introduce a way of constantly forgetting
The results that won't be needed for another 20 years,
At the end of which they'll surface with appropriate loud cheers,
While the ones that won't be needed till forever and a day,
Once their authors get their tenure will be firmly thrown away.

—R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

How likely is it that somebody who solved one of Hilbert's problems is more famous for his work in another part of mathematics? (See p. 304.)

$A_5$ is simple, independent of the five forms in which it commonly arises: as a group of permutations, as the rotation group of a dodecahedron or an icosahedron, as a group of $2 \times 2$ matrices over a field of order 4 or 5 modulo the scalar matrices.

Joseph A. Gallian
Department of Mathematical Sciences
University of Minnesota
Duluth, MN 55812

---

## 145. MISCELLANEA

### Ode to a $\phi$?

A thing as lovely as a $\phi$
I think that I shall never $\left\{ \begin{matrix} \text{see} \\ \text{spy} \end{matrix} \right\}$. [Choose one.]

If you want to make $\phi$ rhyme with "see",
But you're really determined to be
Consistent, admit
Though it doesn't quite fit,
That a disk measures $r^2$ times "pea".

Each of $\xi$, $\pi$, $\phi$, $\chi$, $\psi$ will claim
An identical rhyme for its name.
Before you get hot,
Check Liddell and Scott[1];
Give them, not the author, the blame.

Barry W. Brunson
Department of Mathematics
Western Kentucky University
Bowling Green, KY 42101

[1]A Greek-English Lexicon, compiled by H. G. Liddell & R. Scott, Oxford at the Clarendon Press, 1968 (1st edition 1843;, widely recognized as authoritative).

---

### ANSWER TO PHOTO ON PAGE 269

A. N. Kolmogorov

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## WHEN ONE EQUATION SOLVES THEM ALL

WILLIAM C. WATERHOUSE
*Department of Mathematics, The Pennsylvania State University, University Park, PA 16802*

There is a real polynomial, $X^2 + 1$, with no real root. But as soon as we adjoin a root of this one polynomial, we can solve all other polynomial equations. This fact, the "fundamental theorem of algebra," is a remarkable property of the real numbers, and one would not expect many other number systems to share it. This expectation is actually supported by a theorem, a precise and pretty result due to Artin and Schreier [**1**]. Unfortunately, their theorem is not very widely known. For one thing, it is often buried in more general discussions of ordered fields. But in addition, all the proofs I have seen [**1**], [**2**, p. 47], [**3**, p. 316], [**5**, p. 66] require some special results on extensions of degree $p^2$. In this note, I want to show how the theorem can be proved using only results on extensions of prime degree, results that would be included in any case by most people teaching field theory. I have written out these results as a lemma, so that this note itself can be read by anyone knowing basic Galois theory. Someone interested in the theorem only for fields of characteristic zero can skip parts (2), (4), and (6) of the Lemma and the first and third paragraphs of the main proof.

We call a finite-dimensional field extension *Galois* if it is normal and separable. Our preliminary lemma deals with a Galois field extension $E/F$ of prime degree $p$. In this case the Galois group of automorphisms is cyclic, say with elements $\sigma, \sigma^2, \dots, \sigma^{p-1}, \sigma^p = \mathrm{id}$. For any $x$ in $E$ we define its *norm* and its *trace* by the formulas

$$N(x) = x\sigma(x)\sigma^2(x)\sigma^3(x) \cdots \sigma^{p-1}(x),$$
$$\mathrm{Tr}(x) = x + \sigma(x) + \sigma^2(x) + \cdots + \sigma^{p-1}(x).$$

These elements are obviously sent to themselves by $\sigma$ and thus lie in $F$. Clearly

$$N(xy) = N(x)N(y) \text{ and } \mathrm{Tr}(x+y) = \mathrm{Tr}(x) + \mathrm{Tr}(y),$$

while

$$N(by) = b^p N(y) \text{ and } \mathrm{Tr}(by) = b \cdot \mathrm{Tr}(y)$$

for $b$ in $F$.

LEMMA. *Let $E/F$ be a Galois extension with cyclic Galois group $\langle\sigma\rangle$ of prime order $p$. Then*:

(1) *For any values $c_1, \dots, c_p$ in $E$ not all zero there is some $x$ in $E$ with $0 \neq \sum c_i\sigma^i(x)$.*

(2) $\mathrm{Tr}(E)$ *is all of $F$.*

(3) *Each $z$ in $E$ with $N(z) = 1$ has the form $\sigma(y)/y$ for some $y$ in $E$.*

(4) *Each $z$ in $E$ with $\mathrm{Tr}(z) = 0$ has the form $\sigma(y) - y$ for some $y$ in $E$.*

(5) *If* $\mathrm{char}(F) \neq p$, *and $F$ contains an element $\zeta \neq 1$ with $\zeta^p = 1$, then $E = F(y)$ for some $y$ satisfying $\sigma(y) = \zeta y$ and $y^p \in F$.*

(6) *If* $\mathrm{char}(F) = p$, *then $E = F(y)$ for some $y$ satisfying $\sigma(y) = y + 1$ and $y^p - y \in F$.*

*Proof*. Suppose (1) fails, and let $\sum c_i\sigma^i$ be a counterexample. Clearly there cannot be just one nonzero $c_i$, so say $c_r$ and $c_s$ are nonzero. As $\sigma^r \neq \sigma^s$, we can choose $b$ in $E$ with $\sigma^r(b) \neq \sigma^s(b)$. We have then

$$0 \equiv \sum c_i\sigma^i(bx) \equiv \sum c_i\sigma^i(b)\sigma^i(x),$$

and also

$$0 \equiv \sigma^r(b) \sum c_i \sigma^i(x) \equiv \sum c_i \sigma^r(b) \sigma^i(x).$$

Subtracting, we get

$$0 \equiv \Sigma c_i \left[ \sigma^i(b) - \sigma^r(b) \right] \sigma^i(x).$$

In this relation the coefficient of $\sigma^i$ is zero whenever $c_i$ was zero and is also zero for $i = r$, but it is nonzero for $i = s$. Thus we have a nontrivial counterexample with fewer terms in it. Such reduction cannot continue forever, and hence (1) must be true. Applied to the case $c_i \equiv 1$, this shows that some $\mathrm{Tr}(x)$ is nonzero; and $\mathrm{Tr}(bx) = b\ \mathrm{Tr}(x)$ for every $b$ in $F$, so (2) follows.

For (3), we consider expressions of the form

$$w = x + z\sigma(x) + z\sigma(z)\sigma^2(x) + \cdots + \left(z\sigma(z) \cdots \sigma^{p-2}(z)\right)\sigma^{p-1}(x).$$

By (1), we can choose $x$ to make this nonzero. Applying $\sigma$ and recalling $N(z) = 1$, we find that $z\sigma(w) = w$. Thus we can take $y$ to be $w^{-1}$. For (4) similarly we first use (2) to choose $x$ with $\mathrm{Tr}(x) = -1$. Then we write down

$$y = z\sigma(x) + \left(z + \sigma(z)\right)\sigma^2(x) + \cdots + \left(z + \sigma(z) + \cdots + \sigma^{p-2}(z)\right)\sigma^{p-1}(x).$$

As $\mathrm{Tr}(z) = 0$, we find that $y - \sigma(y)$ equals

$$\begin{aligned} &z\sigma(x) + z\sigma^2(x) + \cdots + z\sigma^{p-1}(x) - \left(\sigma(z) + \cdots + \sigma^{p-1}(z)\right)x \\ &= z\left(x + \sigma(x) + \cdots + \sigma^{p-1}(x)\right) = z\mathrm{Tr}(x) = -z. \end{aligned}$$

For (5), we observe that $N(\zeta) = \zeta^p = 1$, so $\zeta = \sigma(y)/y$ for some $y$. As $\sigma(y) \neq y$, we have $y \notin F$. Hence $E = F(y)$, since the prime degree extension $E/F$ cannot have any intermediate fields. We have

$$\sigma(y^p) = \sigma(y)^p = (\zeta y)^p = \zeta^p y^p = y^p,$$

so $y^p$ is in $F$. For (6) similarly we observe that

$$\mathrm{Tr}(1) = 1 + \cdots + 1 = 0,$$

so $1 = \sigma(y) - y$ for some $y$ by (4). Thus $\sigma(y) = y + 1$. As before, $y$ must generate $E$. Since $(u + v)^p = u^p + v^p$ in characteristic $p$, we have

$$\sigma(y^p - y) = \sigma(y)^p - \sigma(y) = (y + 1)^p - (y + 1) = y^p + 1 - y - 1 = y^p - y,$$

so $y^p - y$ is in $F$. ■

DEFINITION. A field $K$ is called *real closed* if
(1) every element of it is either a square or the negative of a square,
(2) every sum of squares in $K$ is a square,
(3) $-1$ is not a square, and
(4) $K(\sqrt{-1})$ is algebraically closed (that is, every polynomial over it has a root in it).

These conditions, which will all come out automatically when we prove the theorem, obviously say that $K$ is very much like the real numbers. More detailed study ([**4**, §5.6], [**6**, pp. 60–65]) shows in fact that any "elementary" property of the reals is shared by $K$. We may note at least that (2) and (3) imply $\mathrm{char}(K) = 0$, since 1 is a square and $-1$ has the form $1 + \cdots + 1$ in characteristic $p$. Also, (4) implies that every polynomial over $K$ of odd degree has a root in $K$.

THEOREM (Artin-Schreier). *Suppose $K$ is a field whose algebraic closure $\overline{K}$ is a nontrivial finite extension of $K$. Then $K$ is real closed.*

*Proof.* We show first that if $\mathrm{char}(K) = q \neq 0$, then $K$ must be perfect. That is, we show that $K$ is closed under taking $q$th roots. Indeed, $f(x) = x^q$ is a function $\overline{K} \to \overline{K}$ preserving sums as

well as products; its kernel is trivial; and it is surjective, because for any $b$ the equation $X^q - b = 0$ has a root in $\bar{K}$. Thus $f$ is an isomorphism of $\bar{K}$ onto itself. Clearly $f(K) \subseteq K$. If $f(K)$ is a proper subfield of $K$, then we can apply $f^{-1}$ and conclude that $K$ is a proper subfield of $f^{-1}(K)$. Applying $f^{-1}$ again, we see that $f^{-1}(K)$ is a proper subfield of $f^{-2}(K)$, and so on. But such an infinite tower

$$K \subset f^{-1}(K) \subset f^{-2}(K) \subset \cdots \subset \bar{K}$$

cannot exist, since $\bar{K}$ is a finite extension of $K$. Thus $f(K)$ must equal $K$, and $K$ is perfect.

Obviously $\bar{K}$ is a splitting field over $K$, and we have just seen that $K$ is perfect even if $\mathrm{char}(K) \neq 0$, so $\bar{K}/K$ is a Galois field extension. We focus now on some possible intermediate field $F$ for which the degree $|\bar{K}: F|$ is equal to a prime, $p$. Our goal is to restrict the possibilities for this $F$ so strongly that the general result will follow. As $\bar{K}/F$ is Galois of prime degree $p$, we know at least that the results of the lemma can be applied to it.

Suppose first that $\mathrm{char}(F)$ is the same as $p$. Write $\bar{K} = F(y)$ with $y^p - y = a$ in $F$. Choose $b$ in $\bar{K}$ with $\mathrm{Tr}(b) = a$. As $\bar{K}$ is algebraically closed, it contains some $z$ satisfying $z^p - z = b$. Let $c = \mathrm{Tr}(z)$. Then

$$\mathrm{Tr}(z^p) = \sum \sigma^i(z^p) = \sum (\sigma^i(z))^p = \left(\sum \sigma^i(z)\right)^p = (\mathrm{Tr}(z))^p,$$

so $c^p - c = \mathrm{Tr}(z^p - z) = \mathrm{Tr}(b) = a$. Thus $c^p - c = y^p - y$, so

$$c - y = c^p - y^p = (c - y)^p.$$

But the roots of $T^p = T$ are $T = 0, 1, \ldots, p-1$, which all lie in $F$. Thus $c - y$ is in $F$. But $c$ is also in $F$, and hence $y$ is in $F$. This contradicts $\bar{K} = F(y)$. Hence $p \neq \mathrm{char}(F)$.

Suppose next that $p$ is an odd prime different from $\mathrm{char}(F)$. Then the equation

$$0 = (T^p - 1)/(T - 1) = T^{p-1} + \cdots + T + 1$$

does not have 1 as a root. Let $\zeta$ be a root of it (in $\bar{K}$). Since $\bar{K}$ has prime degree $p$ over $F$, there are no fields intermediate between $F$ and $\bar{K}$. Since $\zeta$ satisfies an equation of degree less than $p$, we have $F(\zeta) \neq \bar{K}$, and hence $F(\zeta) = F$. That is, $\zeta$ is in $F$. The roots of $T^p = 1$ then are $1, \zeta, \zeta^2, \ldots, \zeta^{p-1}$, and they are all in $F$. Now by the lemma we can write $\bar{K} = F(y)$ with $\sigma(y) = \zeta y$ and $y^p = a$ in $F$. As $\bar{K}$ is algebraically closed, it contains some $z$ satisfying $z^p = y$. Let $c = N(z)$. Then

$$\begin{aligned} c^p &= (Nz)^p = N(z^p) = N(y) = y(\zeta y)(\zeta^2 y) \cdots (\zeta^{p-1} y) \\ &= y^p \zeta^{p(p-1)/2} = y^p. \end{aligned}$$

Thus $(y/c)^p = 1$, so (as we saw) $y/c$ is in $F$. But $c$ is in $F$, and so $y$ is in $F$. This contradicts $\bar{K} = F(y)$. Hence $p$ cannot be odd.

Thus far we have seen how hard it is for $\bar{K}/F$ to be of prime degree. We continue by finding what restrictions there are in the one case that can occur, $p = 2$ (and $\mathrm{char}(F) \neq 2$). Here by the lemma we can write $\bar{K} = F(y)$ with $\sigma(y) = -y$ and $y^2 = a$ in $F$. In particular, not all elements in $F$ are squares in $F$. If $b$ is any nonsquare in $F$, and $z$ is a fourth root of it in $\bar{K}$, let $c = N(z)$; then

$$c^2 = N(z^2) = N(\sqrt{b}) = -b,$$

so $-b$ is a square in $F$ when $b$ is not. Also, $\sqrt{b} = c\sqrt{-1}$, so $-1$ is not a square in $F$ (and $\bar{K} = F(\sqrt{-1})$). Finally, if $r$ and $s$ are any elements of $F$, then $\bar{K}$ contains some element $u + v\sqrt{-1}$ with

$$(u + v\sqrt{-1})^2 = r + s\sqrt{-1}.$$

This gives

$$r^2 + s^2 = N(r + s\sqrt{-1}) = \left(N(u + v\sqrt{-1})\right)^2,$$

and thus a sum of two squares in $F$ is again a square. In short, $F$ is real closed.

We have shown now that any field $F$ whose algebraic closure has prime degree must be real closed. Consider finally an arbitrary $K$ with $\bar{K}/K$ finite Galois. Let $L = K(\sqrt{-1})$. Then $\bar{K}/L$ is finite Galois. If $L \neq \bar{K}$, choose some nontrivial automorphism $\tau$ of $\bar{K}$ over $L$. Let $n$ be its order, and choose some prime $p$ dividing $n$. Then the field $F$ fixed by $\tau^{n/p}$ has $|\bar{K}: F| = p$. Hence $F$ should be real closed; but this is impossible, since $\sqrt{-1}$ is in $L \subseteq F$. This contradiction shows that $L = \bar{K}$. But then $\bar{K}/K$ has degree 2, and $K$ must be a real closed field. ■

**References**


**1.** E. Artin and O. Schreier, Eine Kennzeichnung der reell abgeschlossenen Körper, Abh. Math. Seminar Hamburg, 5 (1927) 225–231.

**2.** N. Bourbaki, Algèbre VI: Groupes et Corps Ordonnés, Hermann, Paris, 1952.

**3.** N. Jacobson, Lectures in Abstract Algebra III: Theory of Fields and Galois Theory, Van Nostrand, New York, 1964.

**4.** ______, Basic Algebra I, Freeman, San Francisco, 1974.

**5.** I. Kaplansky, Fields and Rings, U. Chicago Press, Chicago, 1969.

**6.** G. Kreisel and J. L. Krivine, Elements of Mathematical Logic, North-Holland, London, 1967.


## THE D'ALEMBERT FUNCTIONAL EQUATION


F. J. PAPP

*Department of Mathematics, University of Michigan, Dearborn, MI 48128*


**I. Introduction.** D'Alembert's functional equation

$$2f(x)f(y) = f(x+y) + f(x-y) \tag{1}$$

has a long history going back to d'Alembert [**1**], Poisson [**2**] and [3], and Picard [**4**] and [**5**, Chapter I, Section III]. The equation plays an important role in determining the sum of two vectors in various Euclidean and non-Euclidean geometries; in addition to the references above see [**6**] and [7, Section 2.4.2].

In equation (1) $f$ is taken to be a continuous real valued function defined on the entire real line. With these assumptions, the only solutions of (1) are

$$f(x) \equiv 0, \quad f(x) \equiv 1, \quad f(x) = \cos kx, \quad \text{and} \quad f(x) = \cosh kx,$$

where $k$ is a real constant. The Cauchy method of showing that these are the only solutions to (1) first determines $f$ on a dense set and then uses continuity to pass to the entire real line; see [**7**, Section 2.4.1] for the details. This paper provides an alternative proof which makes use of a few elementary ideas from calculus and differential equations. This method given below is quite straightforward and will, perhaps, be somewhat more accessible than Cauchy's method.

**II. Preliminary Properties.** We will first obtain several properties which any solution of (1) must satisfy. All of these are derived directly from the functional equation itself.

If $x = y = 0$ in (1), then

$$2[f(0)]^2 = 2f(0) \tag{2}$$

so that, either

$$f(0) = 0 \tag{3}$$

or

$$f(0) = 1. \tag{4}$$

If (3) holds, then for any real $x$,

$$0 = 2f(x)f(0) = f(x+0) + f(x-0) = 2f(x).$$

In other words, $f$ is the identically zero function. So we will assume from now on that (3) does not hold and that $f$ satisfies (4).

Next, any solution of (1) must be an even function. To see this, let $x = 0$ and $y$ be any real number. It then follows from (1) and (4) that

$$2f(y) = 2f(0)f(y) = f(0 + y) + f(0 - y) = f(y) + f(-y);$$

or

$$f(y) = f(-y). \tag{5}$$

**III. Construction of the Continuous Solutions of (1).** Since $f$ is assumed to be continuous on the entire line, it will be integrable on any finite subinterval. Consequently, for $t > 0$ we have

$$\int_{-t}^{t} 2f(x)f(y)\,dy = \int_{-t}^{t} f(x + y)\,dy + \int_{-t}^{t} f(x - y)\,dy. \tag{6}$$

With changes of the variables in the two integrals on the right hand side of this last equation, we obtain

$$2f(x)\int_{-t}^{t} f(y)\,dy = 2\int_{x-t}^{x+t} f(y)\,dy. \tag{7}$$

Since (4) holds, and since we have assumed that $f$ is continuous, there exists $t > 0$ such that

$$\int_{-t}^{t} f(y)\,dy > 0. \tag{8}$$

Note that the right hand side of (7) is differentiable with respect to $x$ and thus so is the left hand side. Knowing now that $f$ is differentiable allows us to conclude that $f''$ exists as well. In fact, proceeding step-by-step, we see that any continuous solution of (1) is in fact infinitely differentiable.

Differentiate both sides of (7) with respect to $x$, evaluate the resulting expression at $x = 0$, and choose $t > 0$ sufficiently small so that (8) holds. It then follows that

$$2f'(0)\int_{-t}^{t} f(y)\,dy = 2[f(t) - f(-t)] = 0, \tag{9}$$

since $f$ is an even function. Consequently

$$f'(0) = 0. \tag{10}$$

Now, differentiate (1) twice with respect to $y$ and obtain

$$2f(x)f''(y) = f''(x + y) + f''(x - y). \tag{11}$$

Letting $y = 0$, we get

$$2f(x)f''(0) = 2f''(x). \tag{12}$$

Let $K = f''(0)$. So (12) becomes (13) below, and we have now shown that if $f$ is a non-trivial solution of (1) it must also be a solution to the initial value problem (13)-(4)-(10):

$$f''(x) = Kf(x), \tag{13}$$

$$f(0) = 1, \tag{4}$$

$$f'(0) = 0. \tag{10}$$

Depending on $K$ there are three possibilities. For $K = 0$,

$$f(x) = C_1 x + C_2;$$

for $K > 0$,

$$f(x) = C_1 \sinh kx + C_2 \cosh kx$$

where $k = K^{1/2}$; and for $K < 0$,

$$f(x) = C_1 \sin kx + C_2 \cos kx$$

with $k = (-K)^{1/2}$.

In all three cases, (4) and (10) lead to $C_1 = 0$ and $C_2 = 1$. Thus the continuous solutions of (1) are

$$f(x) \equiv 0, \tag{14}$$

$$f(x) \equiv 1, \tag{15}$$

$$f(x) = \cosh kx, \tag{16}$$

and

$$f(x) = \cos kx. \tag{17}$$

## References


1. J. d'Alembert, Mémoire sur les principes de mécanique, Hist. Acad. Sci., Paris, 1769, pp. 278–286.
2. S. Poisson, Du parallélogramme des forces, Correspondance sur l'École Polytechnique, I (1804) 356–360.
3. ______, Traité de mécanique, Paris, 1811.
4. C.-E. Picard, Deux leçons sur certaines équations fonctionnelles et la géometrie non-euclidienne, Bull. Soc. Math., XLVI (1922) 404–416, 425–432.
5. ______, Leçons sur quelques équations fonctionnelles avec des applications à divers problèmes d'analyse et de la physique mathématique, Gauthier-Villars, Paris, 1928.
6. R. Bonola, Non-Euclidean Geometry, Dover, New York, 1955.
7. J. Aczél, Lectures on Functional Equations and Their Applications, Academic Press, New York and London, 1966.


# THE DISCRIMINATOR
# (A SIMPLE APPLICATION OF BERTRAND'S POSTULATE)


L. K. Arnold, S. J. Benkoski, and B. J. McCabe

*Daniel H. Wagner, Associates, Station Square One, Paoli, PA 19301*


**1. Introduction.** For a positive integer $n$, what is the smallest value of $k$ such that $1^2, 2^2, \ldots, n^2$ are all incongruent modulo $k$? The answer to this question requires only elementary number theory techniques and also provides a rare application of Bertrand's Postulate in an elementary setting.

Bertrand [**1**] postulated in 1845 that for every natural number $n > 3$, there is a prime $p$ satisfying $n < p < 2n - 2$. He verified this for $n < 3{,}000{,}000$. Tchebychef [**3**] proved this in 1850. Usually, the Postulate is now stated with $n < p < 2n$. Most modern texts use a proof due to Erdős [**2**].

**2. Motivation.** This particular question arose from consideration of a problem in computer simulation. The problem may be described as developing a method to determine quickly the square roots of a long sequence of integers. Each of these integers is a perfect square and the range of the integers is known. However, the integers themselves are generated by another process and, in fact, it can be assumed that they are generated randomly. The goal is to determine the square roots quickly. The usual square root algorithm is much too slow. An alternative procedure would be to calculate all the required square roots once and then "look up" the appropriate value in an array. More precisely, let $S = \{1, 2^2, \ldots, n^2\}$ be the set of integers whose square roots must be determined. Let $A$ be a $1 \times n^2$ array with $A(x) = x^{1/2}$. For any value $s \in S, A(s)$ is the square root of $s$. This is obviously a very quick procedure; however, it requires an array of size $1 \times n^2$. This is a large array and most of the values in the array are not used. Is there an alternative method which would still allow quick calculation of the square roots but require a much smaller

array? A potential candidate is the modulo function. If $1, 2^2, \ldots, n^2$ are distinct modulo $k$, then letting $A(r) = x^{1/2}$ where $r \equiv x \bmod k$ and $1 \leqslant r \leqslant k$, allows a similar look up procedure to be performed. This is obviously slower than the previous method, since the index for that array is determined by reducing $x$ modulo $k$. However, the modulo procedure is, in general, much faster than the square root procedure. The only remaining question is the size of the array required to store the square roots. This is simply the smallest value of $k$ such that $1, 2^2, \ldots, n^2$ are incongruent modulo $k$. This is an interesting number theory problem and the solution is given below.

**3. The discriminator function.** Let $D(n) = \min\{k | 1^2, 2^2, \ldots, n^2$ are incongruent modulo $k\}$. We call $D$ the discriminator (or discriminator function). Table 1 gives the values of $D(n)$ through $n = 18$.

TABLE 1

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $D(n)$ | 1 | 2 | 6 | 9 | 10 | 13 | 14 | 17 | 19 | 22 | 22 | 26 | 26 | 29 | 31 | 34 | 34 | 37 |

We shall show that $D(n)$ (for $n > 4$) is equal to the smallest integer which is greater than or equal to $2n$ and which is either a prime or twice a prime. This is done by first defining $B(n)$. Let

$$B(n) = \begin{cases} 1, & \text{if } n = 1, \\ 2, & \text{if } n = 2, \\ 6, & \text{if } n = 3 \\ 9, & \text{if } n = 4, \\ \min\{k | k \geqslant 2n \text{ and } k = p \text{ or } k = 2p \text{ for } p \text{ a prime}\}, & \text{if } n > 4. \end{cases}$$

We show below that $D(n) = B(n)$ by first showing that $D(n) \leqslant B(n)$ and then that $B(n) \leqslant D(n)$. (Throughout it is assumed that $n > 4$.)

The proof that $D(n) \leqslant B(n)$ is done by two lemmas.

LEMMA 1. *If $p \geqslant 2n$, then $D(n) \leqslant p$.*

*Proof.* Suppose not. $D(n) > p$ implies that there exist $s$ and $r$ with $1 \leqslant r < s \leqslant n$ such that $s^2 \equiv r^2 \bmod p$. This implies that $p | (s + r)(s - r)$. This is impossible since $p \geqslant 2n$ implies that $p > (s + r)$.

LEMMA 2. *If $2p \geqslant 2n$, then $D(n) \leqslant 2p$.*

*Proof.* Suppose not. $D(n) > 2p$ implies there exist $r$ and $s$ with $1 \leqslant r < s \leqslant n$ such that $s^2 \equiv r^2 \bmod 2p$. Thus, $2p | (s + r)(s - r)$. Since $s - r < p$, then $p | (s + r)$. Also, $2p \geqslant 2n > s + r$ implies that $p = r + s$. Thus, $2 | (s - r)$. But this means that $s + r$ is odd and $s - r$ is even.

These two lemmas show that $D(n) \leqslant B(n)$. Next we show that $B(n) \leqslant D(n)$. This is also done with two lemmas.

LEMMA 3. $D(n) \geqslant 2n$.

*Proof.* If not, then there exist integers $r$ and $s$ such that $1 \leqslant r < s \leqslant n$ and $D(n) = s + r$. This is impossible since $s^2 \equiv r^2 \bmod r + s$.

A check of Table 1 indicates that $B(n) = D(n)$ for $n \leqslant 18$. The following lemma completes the proof.

LEMMA 4. *If $2n \leqslant m < B(n)$ and $n > 18$, then there exist integers $s$ and $r$ such that $1 \leqslant r < s \leqslant n$ and $s^2 \equiv r^2 \bmod m$.*

*Proof.* Since $m < B(n)$, this means that $m$ is less than the smallest prime greater than or equal

to $2n$. Bertrand's Postulate states that there is at least one prime between $x$ and $2x$. This means that $m < 4n$. Further, $m < B(n)$ implies that $m \neq p$ and $m \neq 2p$ for any prime $p$. Thus, writing $m$ as a product of primes

$$m = 2^{\alpha_1}3^{\alpha_2}5^{\alpha_3}\cdots,$$

we see that it can be factored in at least one of four ways:

(1) $\alpha_1 = 1$ and $m = 2t^2$.
(2) $\alpha_1 = 1$ and $m = 2ab$ with $a \geqslant 3$, $b \geqslant 3$, and $a$ and $b$ both odd.
(3) $\alpha_1 \neq 1$ and $m = t^2$.
(4) $\alpha_1 \neq 1$ and $m = ab$ with $a \geqslant b \geqslant 2$, and $a$ and $b$ of the same parity.

In each case we will exhibit $s$ and $r$ such that $1 < r < s$, $s^2 \equiv r^2 \bmod m$, and $s \leqslant n$. (We will use the fact that for $a, b, s$ positive integers with $a, b \geqslant s$, then $(a+b) \leqslant ((ab/s)+s)$. This follows immediately from $s(a-s) \leqslant b(a-s)$.)

(1) Let $s = 3t$ and $r = t$, then $s^2 \equiv r^2 \bmod m$. Since $2t^2 < 4n$, this implies that $t < (2n)^{1/2}$ or $s < (18m)^{1/2}$. Now, $(18n)^{1/2}$ is less than $n$ for $n \geqslant 18$. Thus, $s \leqslant n$.

(2) Let $s = a + b$ and $r = a - b$, then $s^2 \equiv r^2 \bmod m$. For $n > 9$, it follows that $((4n/6)+3) < n$. Since $2ab < 4n$, then $((ab/3)+3) < ((4n/6)+3) < n$. Combining this with the remark above yields $s = a + b \leqslant ((ab/3)+3) < n$.

(3) Let $s = 2t$ and $r = t$. Clearly, $s^2 \equiv r^2 \bmod m$. For $n \geqslant 16$, $s = 2t < 4n^{1/2} \leqslant n$.

(4) Let $s = (a+b)/2$ and $r = (a-b)/2$. Again, $s^2 \equiv r^2 \bmod m$. From the remark above, $(a+b) \leqslant ((ab/2)+2)$. This implies that

$$(a+b)/2 = s \leqslant (ab/4+1) = (m/4+1).$$

Since $m < 4n$, we have $s < (4n/4+1) = n+1$. Thus, $s \leqslant n$.

**4. Summary.** If $D(n)$ denotes the smallest positive integer $k$ such that $1^2, 2^2, \ldots, n^2$ are incongruent modulo $k$, then

$$D(n) = \begin{cases} 1, & \text{if } n = 1, \\ 2, & \text{if } n = 2, \\ 6, & \text{if } n = 3, \\ 9, & \text{if } n = 4, \\ \min\{k \mid k \geqslant 2n \text{ and } k = p \text{ or } k = 2p \text{ for } p \text{ a prime}\}, & \text{if } n > 4. \end{cases}$$

**References**

1. J. Bertrand, Mémoire sur le nombre de valeurs que peut prendre une fonction quand on y permute les lettres qu'elle renferme, J. Ec. Polyt., 30 (1845) 123–140.
2. P. Erdős, Beweiss eines Satzes von Tschebyschef, Acta Litt. Sci. Reg. Univ. Hungar, Fr.-Jos., Sect. Sci. Math., 5 (1932) 194–198.
3. P. L. Tchebychef, Mémoire sur les nombres premiers, St. Pet. Ac. Mm., VII (1854) 17–33.

## NOTE ON GRAPHS AND MATRIX INEQUALITIES

SEYMOUR LIPSCHUTZ
*Department of Mathematics, Temple University, Philadelphia, PA 19122*

**1. Introduction.** The purpose of this note is to show how a very simple counting argument in graph theory can be used to obtain results about matrices which are related to a classical problem in analysis. (We use the notation $S(M)$ to denote the sum of the entries of a matrix $M$.)

**2. Counting Argument.** Let $G$ be a finite graph, let $P_G(k)$ denote the number of paths of

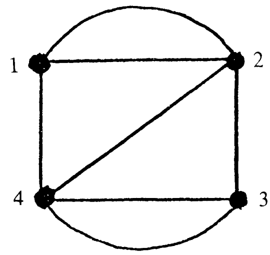



FIG. 1

length $k$ in $G$, and let $d_i$ denote the degree of the vertex $v_i$ of $G$. We illustrate our idea by means of an example. Suppose $G$ is the graph in Fig. 1. The adjacency matrix $A$ of $G$, and its square $A^2$ follow.

$$A = \begin{pmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 2 \\ 1 & 1 & 2 & 0 \end{pmatrix} \quad \text{and} \quad A^2 = \begin{pmatrix} 5 & 1 & 4 & 2 \\ 1 & 6 & 2 & 4 \\ 4 & 2 & 5 & 1 \\ 2 & 4 & 1 & 6 \end{pmatrix}.$$

Then it is well known that the number $P_G(2)$ of paths of length two in $G$ is obtained by summing the elements of $A^2$:

$$P_G(2) = S(A^2) = 12 + 13 + 12 + 13 = 50.$$

(We compute $S(A^2)$ by first adding the elements of each row.)

Our idea is to count $P_G(2)$ in another way. First observe that the degrees of the vertices of $G$ are as follows:

$$d_1 = 3, \qquad d_2 = 4, \qquad d_3 = 3, \qquad d_4 = 4.$$

Now each path of length two in $G$ has a unique midpoint. The number of paths of length two with midpoint $v_1$ of $G$ is equal to $d_1^2 = 3^2 = 9$, since there are exactly $d_1 = 3$ ways to enter $v_1$ and $d_1 = 3$ ways to leave $v_1$. A similar result holds for the other vertices of $G$. Accordingly,

$$P_G(2) = d_1^2 + d_2^2 + d_3^2 + d_4^2 = 9 + 16 + 9 + 16 = 50.$$

This result clearly holds for any finite graph.

PROPOSITION 1. *Let $G$ be a finite graph with adjacency matrix $A$. Then*

$$P_G(2) = S(A^2) = d_1^2 + d_2^2 + \cdots + d_n^2,$$

*where $d_i$ denotes the degree of the vertex $v_i$ of $G$.*

**3. Classical Problem.** Suppose $A$ is a nonnegative irreducible $n$-square matrix. Then it is well known (see, for example, [**1**], p. 27) that $A$ has a positive eigenvalue $r(A)$, called the Perron root of $A$, such that

$$|\lambda| \leqslant r(A)$$

for any other eigenvalue $\lambda$ of $A$. A great deal of research has been done on finding bounds for $r(A)$. (See, for example, [**2**] and [**3**].) When $A$ is a nonnegative symmetric matrix, the inequality

$$S(A)/S(I) \leqslant S(A^3)/S(A^2)$$

is needed to obtain certain bounds for the Perron root of $A$. (Here $I$ denotes the identity matrix.) This inequality, however, does not hold for all nonnegative symmetric matrices $A$ of order greater than three [**4**]. Using Proposition 1, we easily prove Theorem 1.

THEOREM 1. *Let $A$ be a nonnegative symmetric matrix. Then*

$$S(A)/S(I) \leqslant S(A^2)/S(A).$$

*Proof*. We need only prove the theorem when $A$ is an integral matrix; it follows that the inequality will hold for rational entries and, by passing to the limit, the inequality will also hold for real entries.

Let $G$ be the multigraph whose adjacency matrix is $A$. Since $d_i$ denotes the degree of the vertex $v_i$, it is clear that $d_i$ is also equal to the sum of the elements of row $i$ of $A$. Hence

$$S(A) = d_1 + d_2 + \cdots + d_n.$$

The Cauchy-Schwartz inequality with $(1,1,\ldots,1)$ and $(d_1, d_2, \ldots, d_n)$ gives

$$(d_1 + d_2 + \cdots + d_n)^2 \leqslant n(d_1^2 + d_2^2 + \cdots + d_n^2).$$

But $S(I) = n$. Therefore, by Proposition 1,

$$(S(A))^2 \leqslant S(I)S(A^2)$$

which gives our result.

This theorem has already been proved in [**5**]. Our proof, however, is geometrical using Proposition 1. Later we prove Theorem 1 without the restriction that $A$ be nonnegative.

**4. Generalization of Counting Argument.** Again let $G$ be a multigraph with $n$ vertices and let $A = (a_{ij})$ be its adjacency matrix. Let

$$d_i(k) = \text{number of paths of length } k \text{ beginning at vertex } v_i.$$

(Note $d_i(1)$ equals $d_i$, the degree of $v_i$.) Then it is well known that

$$d_i(k) = \text{sum of the elements of row } i \text{ of the matrix } A^k.$$

Hence

$$S(A^k) = d_1(k) + d_2(k) + \cdots + d_n(k).$$

Consider now the number $P_G(2k)$ of paths of length $2k$ in $G$. It is well known that $P_G(2k)$ is equal to the sum of the elements of the matrix $A^{2k}$. On the other hand, as with paths of length two, each path of length $2k$ has a midpoint and the number of paths of length $2k$ with midpoint $v_i$ is precisely $d_i(k)^2$. Thus we have the following generalization of Proposition 1.

PROPOSITION 2. *Let $G$ be a finite graph with adjacency matrix $A$. Then*

$$P_G(2k) = S(A^{2k}) = d_1(k)^2 + d_2(k)^2 + \cdots + d_n(k)^2,$$

*where $d_i(k)$ is the number of paths of length $k$ beginning at vertex $v_i$.*

**5. Recursion Formula and Theorem 2.** Consider now $d_i(k+1)$, the number of paths of length $k+1$ beginning at vertex $v_i$. Then $a_{ij}d_j(k)$ is equal to the number of paths of length $k+1$ whose first vertex is $v_i$ and whose second vertex is $v_j$. Hence the sum of these numbers for all the vertices $v_j$ gives us $d_i(k+1)$.

PROPOSITION 3. *Let $G$ be a graph with adjacency matrix $A = (a_{ij})$. Then*

$$d_i(k+1) = \sum_j a_{ij} \cdot d_j(k).$$

We are now able to prove the following theorem.

THEOREM 2. *Let $A$ be any nonnegative symmetric matrix. Then for any positive integer $k$,*

$$S(A^{2k+2}) \leqslant c(A)S(A^{2k}),$$

*where $c(A)$ is the sum of the squares of the elements of $A$.*

*Proof*. As with Theorem 1, we need only prove the theorem when $A$ is an integral matrix. By Proposition 2 and Proposition 3,

$$S(A^{2k+2}) = \sum_i d_i(k+1)^2 = \sum_i \Big(\sum_j a_{ij} d_j(k)\Big)^2.$$

By the Cauchy-Schwartz inequality and Proposition 2, for each $i$,

$$\Big(\sum_j a_{ij} d_j(k)\Big)^2 \leqslant \Big(\sum_j a_{ij}^2\Big)\Big(\sum_j d_j(k)^2\Big) = \Big(\sum_j a_{ij}^2\Big) S(A^{2k}).$$

Therefore

$$S(A^{2k+2}) \leqslant \sum_i \Big(\sum_j a_{ij}^2\Big) S(A^{2k}) = \Big(\sum_{i,j} a_{ij}^2\Big) S(A^{2k}) = c(A) S(A^{2k}).$$

Thus the theorem is proved. (Observe that $c(A)$ does not depend on $k$.)

**6. Generalization of Theorem 1.** The restriction that $A$ be nonnegative in Theorem 1 is not necessary.

THEOREM 3. *Let $A$ be a symmetric matrix. Then*

$$S(A)/S(I) \leqslant S(A^2)/S(A).$$

*Proof.* Let $d_i$ denote the sum of the elements of row $i$ of $A$. Then

$$S(A) = d_1 + d_2 + \cdots + d_n.$$

Using $A = (a_{ij})$, we have

$$S(A^2) = \sum_i \sum_j \Big(\sum_k a_{ik} a_{kj}\Big).$$

Using $a_{ik} = a_{ki}$, and interchanging the summands, we have

$$S(A^2) = \sum_k \Big(\sum_i \sum_j a_{ki} a_{kj}\Big) = \sum_k \Big(\sum_i a_{ki}\Big)\Big(\sum_j a_{kj}\Big).$$

But

$$\sum_i a_{ki} = d_k \text{ and } \sum_j a_{kj} = d_k.$$

Hence

$$S(A^2) = \sum_k d_k d_k = d_1^2 + d_2^2 + \cdots + d_n^2.$$

The theorem now follows from the Cauchy-Schwartz inequality as in Theorem 1.

REMARK. Although the proof is actually only a few lines, it lacks the geometrical interpretation of the term $d_k^2$.

**References**

1. A. Berman and R. J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, Academic Press, New York, 1979.

2. A. Brauer and I. C. Gentry, Bounds for the greatest characteristic root of an irreducible nonnegative matrix: II, Linear Algebra and Appl., 13 (1976) 109–114.

3. E. Deutsch, Bounds for the Perron root of a nonnegative irreducible partitioned matrix, Pacific J. Math., 92 (1981) 49–56.

4. D. London, Two inequalities in nonnegative symmetric matrices, Pacific J. Math., 16 (1966) 515–536.

5. M. Marcus and M. Newman, The sum of elements of the powers of a matrix, Pacific J. Math., 12 (1962) 627–635.

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## WIENER CHAINS

KURT KREITH
*Department of Mathematics, University of California, Davis, CA 95616*

**1. Introduction.** The study of Markov chains constitutes a useful and important enterprise which deserves to be pursued for its own sake. However, for many students it also represents essential preparation for a profound and challenging generalization—the study of stochastic differential equations.

This note deals with an exercise which, if presented in the context of stationary Markov chains, should also prepare the student to understand some of the basic ideas underlying continuous processes whose randomness is the result of "white noise". It is based on a rather obvious discrete analogue of the Wiener process, one whose properties will not surprise anyone familiar with the continuous version. Indeed, the only remarkable thing about the ideas which follow is that they are not embodied among the standard exercises of introductory texts dealing with Markov chains [**1**], [**2**], [**3**].

**2. Some Motivational Experiments.** Motivated by the classical "gambler's ruin" situation, we ask the student to consider two individuals who agree to toss a fair coin until the difference between the total number of heads and tails exceeds some large number, say $d$. Instead of calculating the distribution corresponding to this difference directly, we proceed empirically by representing the resulting stationary Markov chain in terms of a $(2d+1)\times(2d+1)$ transition matrix $P$ with transition probabilities

$$P(x,x+1)=P(x,x-1)=\frac{1}{2} \tag{2.1}$$

for all $x\in\mathscr{S}=\{-d,-d+1,\ldots,0,1,\ldots,d\}$, except that $P(-d,-d)=P(d,d)=1$. Adopting the notation [**2**]

$$\pi_n(x)=\Pr(X_n=x);\qquad \vec{\pi}_n=(\pi_n(-d),\ldots,\pi_n(0),\ldots,\pi_n(d)),$$

we readily verify by several applications of

$$\vec{\pi}_{n+1}=\vec{\pi}_nP \tag{2.2}$$

that an initial state $X_0$ with $\pi_0(0)=1$ leads to $X_n$ whose distributions involve binomial coefficients. That is,

$$\vec{\pi}_1=\left(0,\ldots,0,0,\frac{1}{2},0,\frac{1}{2},0,0,\ldots,0\right),$$

$$\vec{\pi}_2=\left(0,\ldots,0,\frac{1}{4},0,\frac{1}{2},0,\frac{1}{4},0,\ldots,0\right),$$

$$\vec{\pi}_3=\left(0,\ldots,\frac{1}{8},0,\frac{3}{8},0,\frac{3}{8},0,\frac{1}{8},\ldots,0\right),\text{ etc.}$$

The reason for this pattern can of course be determined by showing that as long as $n\leqslant d$, $X_n=2Y_n-n$, where $Y_n$ is binomial with mean $n/2$ and variance $n/4$. This provides the theoretical explanation (which might best be suppressed for the moment) underlying the empirical observation that the $X_n$ generated by this process have distributions with "binomial patterns" and

that they satisfy

$$E(X_n) = 0; \qquad \operatorname{Var}(X_n) = n$$

for $n = 0, 1, \dots, d$.

To eliminate the zeros which separate the binomially patterned states in the above $\vec{\pi}_i$ (i.e., to deal with the periodicity of this particular chain) one can modify (2.1) so that

$$P(x, x+1) = P(x, x-1) = \frac{1}{4}, \qquad P(x, x) = \frac{1}{2}, \tag{2.3}$$

except for the special cases of $P(-d, -d) = P(d, d) = 1$. Starting with $\pi_0(0) = 1$, one again obtains $X_n$ whose probability functions give rise to binomial patterns symmetric about $x = 0$. Specifically, we now find that

$$\vec{\pi}_1 = \left(0, \dots, 0, 0, \frac{1}{4}, \frac{2}{4}, \frac{1}{4}, 0, 0, \dots, 0\right),$$

$$\vec{\pi}_2 = \left(0, \dots, 0, \frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}, 0, \dots, 0\right),$$

$$\vec{\pi}_3 = \left(0, \dots, \frac{1}{64}, \frac{6}{64}, \frac{15}{64}, \frac{20}{64}, \frac{15}{64}, \frac{6}{64}, \frac{1}{64}, \dots, 0\right), \text{ etc.,}$$

for which $E(X_n) = 0$ and the variance grows with $n$ according to the rule $\operatorname{Var}(X_n) = n/2$ (as long as $n < d$). As before, there is a simple theoretical explanation—i.e., that $X_n + n$ is binomial with parameters $2n$ and $\frac{1}{2}$.

The point of such experimentation is to get a feel for the special role which binomially-related distributions play in certain Markov processes. While they do not constitute stationary distributions, they do "reproduce", in the sense that if $\vec{\pi}_n$ is the probability vector associated with an appropriately translated binomial $X_n$, then $\vec{\pi}_{n+1} = \vec{\pi}_n P$ will contain the same binomial pattern, but with a larger variance.

The next problem is to gain some insight into the kinds of Markov processes which serve to reproduce binomially-related distributions in this way. Such insight follows readily from the relation

$$\vec{\pi}_{m+n} = \vec{\pi}_m P^n \tag{2.4}$$

which shows that powers of the matrices determined by (2.1) or (2.3) also have reproducing properties similar to $P$. To find the form of such $P^n$ it is *not* necessary to compute matrix products. Rather, choosing $\vec{\pi}_0$ with $\pi_0(y) = \delta_{xy}$, the relation

$$\vec{\pi}_m = \vec{\pi}_0 P^m \tag{2.5}$$

shows that the $x$th row of $P^m$ is just $\vec{\pi}_m$. Thus it is transition matrices whose $x$th rows are symmetrically distributed about $y = x$ in a binomial pattern which have the previously observed reproducing property vis-a-vis binomial distributions. All this is based, of course, on the assumption that $d$ is sufficiently large so that the end conditions $P(-d, -d) = P(d, d) = 1$ do not enter into the matrix computations.

**3. The Formalism.** At this point the student should be prepared to let $d \to \infty$ and to accept the following definitions.

DEFINITION 1. A random variable $X$ is *$n\sigma^2/2$-centered binomial* (with mean 0 and variance $n\sigma^2/2$) if $(X_n + n)/\sigma$ is binomial with parameters $2n$ and $\frac{1}{2}$.

Thus the random variables $X_1, X_2, \dots$ generated by (2.3) and $X_0 = 0$ have now been dubbed "$n/2$-centered binomial". Motivated by the process observed under assumptions (2.3) we go on to define a *Wiener chain* as a stationary Markov chain $W_0, W_1, W_2, \dots$ whose state space consists of

all integers and for which

(3.1)
(i) $W_0 = 0$,
(ii) $W_n$ is $n\sigma^2/2$-centered binomial,
(iii) If $n_1 < n_2 < n_3 < n_4 < \cdots$, then $W_{n_2} - W_{n_1}, W_{n_3} - W_{n_2}, W_{n_4} - W_{n_3}, \ldots$ are mutually independent.

This definition has an important consequence.

THEOREM 3.1. *If $W_0, W_1, \ldots$ is a Wiener chain, then $W_{m+n} - W_m$ is $n\sigma^2/2$-centered binomial.*

*Proof.* Because $W_0, W_1, \ldots$ is assumed to be a stationary Markov chain, we have

$$\begin{aligned}
\Pr(W_{m+n} - W_m = x) &= \sum_y \Pr(W_{m+n} = x + y \text{ and } W_m = y)\\
&= \sum_y \Pr(W_{m+n} = x + y | W_m = y)\Pr(W_m = y)\\
&= \sum_y \Pr(W_{m+n} = x | W_m = 0)\Pr(W_m = y)\\
&= \Pr(W_{m+n} = x | W_m = 0)\\
&= \Pr(W_n = x | W_0 = 0).
\end{aligned}$$

However, since $W_0 = 0$, the last sum reduces to $\Pr(W_n = x)$, as was to be proved.

It is also easy to verify that the transition probabilities giving rise to a $n/2$-centered binomial Wiener chain are just those of (2.3):

$$P(x, x+1) = P(x, x-1) = \frac{1}{4}, \qquad P(x, x) = \frac{1}{2}. \tag{3.2}$$

More generally, if $k$ is a positive integer we can find the transition probabilities $P(x, y)$ which, together with $W_0 = 0$, give rise to a $nk^2/2$-Wiener chain. Specifically $P(x, y)$ is the probability function of the random variable $Y_x$, where $(Y_x + x + n)/k$ is binomial with parameters $2n$ and $\frac{1}{2}$.

This structure provides the basis for a variety of exercises which illustrate, in a discrete context, the phenomena central to "white noise" processes—e.g., diffusion, semigroups, and reproducing distributions.

**4. An Application.** As is the case with most aleatory processes, the central ideas related to Wiener chains are well illustrated in a gambling context. Consider a conservative investor who, while willing to take substantial risks, also seeks a complete mathematical description of the uncertainties he faces. With initial capital $X_0$ and an income of $b_n$ at the end of the $n$th month, he plans to invest in a venture whose monthly gains and losses are described in terms of a random variable $Y_n$. He is able to vary his monthly commitment to the investment fund by means of a multiplicative constant $a_n \geqslant 0$. Thus his financial situation can be described by means of a stochastic difference equation

$$X_n = X_{n-1} + b_n + a_n(Y_n - Y_{n-1}) \tag{4.1}$$

for which $X_0$ is given and $Y_0 = 0$.

Such difference equations are readily summed to yield

$$X_n = X_0 + \sum_{j=1}^{n} b_j + \sum_{j=1}^{n} a_j(Y_j - Y_{j-1}). \tag{4.2}$$

Assuming that the investment venture is "fair" in the sense that $E(Y_j) = 0$ for $j = 1, 2, \ldots,$ we can take expectations of (4.2) to establish the intuitively obvious fact that

$$E(X_n) = X_0 + \sum_{j=1}^{n} b_j. \tag{4.3}$$

However, the questions which really interest our investor are not answered by (4.3). Concerns such as, "What are the chances that I will go broke?", or hopes such as, "Can I become a millionaire in two years?", require more information about the distribution of $X_n$. For most investment schemes such information is very hard to come by, but there are some special forms of stochastic difference equations for which we can obtain an explicit representation of $X_n$.

For example, it is readily shown from (4.2) and (4.3) that if $E(Y_j) = 0$ for $j = 1, 2, \ldots,$ then

$$\operatorname{Var}(X_n) = E\left(\left[\sum_{j=1}^{n} a_j(Y_j - Y_{j-1})\right]^2\right). \tag{4.4}$$

However, in most situations the computation of the right side of (4.4) poses a major challenge.

The point which was made by our experiments of Section 2 is that the situation is very different if one invests in a $nk^2/2$-Wienerspiel—i.e., an investment program whose payoffs are given by $W_n$ instead of $Y_n$. For now the equation

$$X_n = X_0 + \sum_{j=1}^{n} b_j + \sum_{j=1}^{n} a_j(W_j - W_{j-1}), \tag{4.5}$$

combined with the fact that the sums of independent binomial random variables are again binomial, shows that $X_n$ will itself be a shifted binomial random variable whose distribution is fully determined by the mean and variance. Since the mean of $X_n$ is given by (4.3), it is straightforward to use (4.4) and the independence of the various $\Delta W_j = W_j - W_{j-1}$ to determine that

$$\begin{aligned}\operatorname{Var}(X_n) &= \sum_{j=1}^{n} a_j^2 E(W_j - W_{j-1})^2 \\ &= \frac{k^2}{2}\sum_{j=1}^{n} a_j^2,\end{aligned}$$

thereby providing a complete mathematical description of the process.

**5. Epilogue.** It only remains to rewrite (4.1) in the form

$$\Delta X_n = b_n + a_n \Delta Y_n \tag{5.1}$$

and to ponder a passage to the limit as investments are compounded at increasingly shorter time intervals. One is tempted to represent the limiting process in the form

$$dX = b(t) + a(t)\, dY, \qquad X(0) = X_0 \tag{5.2}$$

and to pose analogous questions in this continuous case. For a very special class of $dY$—i.e., for $dY = dW$, when $dW$ is a form of "white noise"—equation (5.2) also allows a complete solution in terms of normal, rather than binomial, distributions.

But this is another story, one which this note was intended to motivate rather than tell. A very readable introduction to continuous Wiener processes is given in [**2**].

### References

**1.** K. L. Chung, Markov Chains with Stationary Transition Probabilities, Springer-Verlag, Berlin, 1960.

**2.** Hoel, Port, and Stone, Introduction to Stochastic Processes, Houghton Mifflin, Boston, 1972.

**3.** J. Kemeny and L. Snell, Finite Markov Chains, Van Nostrand, Princeton, 1960.

## THE BISECTION METHOD: A BEST CASE ANALYSIS

A. FINBOW
*Department of Mathematics, Saint Mary's University, Halifax, Canada B3H 3C3*

Let $p$ be a polynomial of degree $n$ with integer coefficients and no rational roots in the interval $[a, b]$, where sign $(p(a)) \neq$ sign $(p(b))$ and $a$ and $b$ are rational. Then we are in a position to use the bisection method to approximate a root of $p$ lying in this interval. In this note we compute a uniform (in $p$) lower bound on the number of iterations required of this method to obtain an $\varepsilon$-root for $p$ (i.e., a point $x$ such that $|p(x)| < \varepsilon$.)

The bisection method for approximating roots is usually introduced in a first course in Computing Science and works for any continuous real function $f$ of a real variable. The algorithm is recursive and the $(k+1)$st step $(k > 0)$ requires of its predecessor an interval $[a_k, b_k]$ with sign $(f(a_k)) \neq$ sign $(f(b_k))$, and an approximate root $r_k$ lying in this interval. It then proceeds as follows: if $r_k$ is a root, the algorithm terminates; otherwise a new interval and approximate root are computed by

$$[a_{k+1}, b_{k+1}] = \begin{cases} [r_k, b_k], & \text{if sign}\,(f(r_k)) = \text{sign}\,(f(a_k)), \\ [a_k, r_k], & \text{otherwise,} \end{cases}$$

and

$$r_{k+1} = (a_{k+1} + b_{k+1})/2.$$

The user must seed the algorithm with an initial interval $[a, b]$ with sign $(f(a)) \neq$ sign $(f(b))$ at which point (the first step) the assignments $a_1 = a$, $b_1 = b$ and $r_1 = (a + b)/2$ are made.

The intermediate value theorem implies that the approximation $r_k$ is within $(b-a)/2^k$ of a root regardless of $f$. If, however, one requires that $r_k$ be an $\varepsilon$-root, then an upper bound on the number of iterations depends upon the function. (If, for example, $f$ is differentiable and a bound $M$ on $\{|f'(x)|\colon x \in [a, b]\}$ is known, then $|f(r_k)| \leqslant (b-a)M/2^k$.) The same is true in general of a lower bound, but we can do better when an irrational root of a polynomial with integer coefficients is being approximated on an interval with rational endpoints:

THEOREM. *Let $p(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0$ be a polynomial with integer coefficients and no rational roots in $[a, b]$, where* sign $(p(a)) \neq$ sign $(p(b))$ *and where $a = d/e$ and $b = g/h$ for some integers $d, e, g, h$ with $e > 0$, $h > 0$. Then, if $\{r_k\}$ is the sequence of approximate roots obtained by applying the bisection method to $p$ in the interval $[a, b]$, a necessary condition for $r_k$ to be an $\varepsilon$-root is*

$$k > \frac{1}{n}\log_2(\varepsilon e^n h^n)^{-1}.$$

*Proof*. Note that for each $k$, $r_k = s_k/t_k$ for some integer $s_k$ and $t_k = 2^k eh$. Now since $r_k$ is rational, we can conclude that

$$0 \neq |p(r_k)| = \frac{|c_n s_k^n + c_{n-1}s_k^{n-1}t_k + \cdots + c_0 t_k^n|}{t_k^n}.$$

The numerator of this fraction is thus a positive integer and thence at least 1. Therefore, we conclude that

$$|p(r_k)| \geqslant t_k^{-n} = (eh2^k)^{-n},$$

from which our result follows.

We remark that the hypothesis that $p$ have no rational roots in $[a, b]$ can be weakened to say that $p$ have no roots in $[a, b]$ of the form $r = a + (b-a)i/2^j$ for integers $i$ and $j$, since this

weakened hypothesis is equivalent to nontermination of the algorithm (cf. "Linear Convergence and the Bisection Algorithm" by Edwin H. Kaufman, Jr., and Terry D. Lenker, submitted to this Journal).

Finally, we remark that there is no such polynomial-independent lower bound on the number of iterations $k$ required in order that $r_k$ be within distance $\varepsilon$ of a root. Indeed, pick an integer $M$ such that $\sqrt{M} > (b-a)/2$ and $\sqrt{M}$ is irrational. Then the family of quadratic polynomials

$$\{c^2x^2 - 2cdx + d^2 - c^2M \colon c \text{ and } d \text{ are integers with } c \neq 0\}$$

has the property that (a) the set of roots of its elements is a dense subset of the irrationals, while (b) each of its elements has at most one root in $[a, b]$.

---

# PROBLEMS AND SOLUTIONS

EDITED BY G. L. ALEXANDERSON, DAVID BORWEIN (ADVANCED PROBLEMS), H. M. W. EDGAR (ELEMENTARY PROBLEMS), AND D. H. MUGLER



*Send all* ***proposed*** *problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions*** *should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by August 31, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3083. *Proposed by Gunnar Blom, University of Lund and Lund Institute of Technology, Sweden.*

If the relation

$$x^3 + y^3 + z^3 = (x+y)(x+z)(y+z)$$

is satisfied by $x_0, y_0, z_0$, where

$$(x_0^2 + y_0^2 - z_0^2)(x_0^2 + z_0^2 - y_0^2)(y_0^2 + z_0^2 - x_0^2) \neq 0,$$

then it is also satisfied by

$$x_1 = \frac{1}{y_0^2 + z_0^2 - x_0^2}, \qquad y_1 = \frac{1}{x_0^2 + z_0^2 - y_0^2}, \qquad z_1 = \frac{1}{x_0^2 + y_0^2 - z_0^2}.$$

E 3084. *Proposed by Paris Pamfilos, University of Crete, Greece.*

Given a family of circles in the plane all of which pass through a common point and no two of which are equal, show that there is another circle enveloping all the circles of the family if and only if there is a straight line containing all the intersection points of the common tangents of any two circles of the family.

E 3085. *Proposed by T. C. Lim, George Mason University, Fairfax, VA.*

Let $g(\mu)$ be the unique nonnegative solution of

$$\{\mu + g(\mu)\}^p + |\mu - g(\mu)|^p = 2\mu,$$

where $1 < p < 2$ and $0 \leqslant \mu \leqslant \frac{1}{2}$. Prove that

$$\{1 - \mu + g(\mu)\}^p + |1 - \mu - g(\mu)|^p \leqslant 2(1 - \mu).$$

E 3086. *Proposed by Dennis Spellman, Sacred Heart University, Bridgeport, CT.*

If $c$ and $m$ are positive integers each greater than 1, find the number $n(c, m)$ of ordered $c$-tuples $(n_1, n_2, \ldots, n_c)$ with entries from the initial segment $\{1, 2, \ldots, m\}$ of the positive integers such that $n_2 < n_1$ and $n_2 \leqslant n_3 \leqslant \cdots \leqslant n_c$.

E 3087. *Proposed by Weixuan Li, University of Waterloo, and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Let $a_i$, $i = 1, 2, \ldots, n$, be real numbers such that $0 \leqslant a_i \leqslant 1$, where $n \geqslant 2$. Find a best upper bound for

$$S_n = (a_1 - a_2)^2 + (a_2 - a_3)^2 + \cdots + (a_n - a_1)^2,$$

and determine all cases when this bound is attained.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Summing Square Roots

E 2941 [1982, 273]. *Proposed by Richard Johnsonbaugh, Chicago State University.*

In L. W. Gates, *Summing square roots*, Math. Gaz., 64, No. 428, June 1980, pp. 86–89,

$$S_n = \frac{1}{24}\left[(8n + 5)(4n + 1)^{1/2} - 5\right]$$

is proposed as an approximation to

$$G_n = \sum_{t=1}^{n} t^{1/2}.$$

Prove that $\lim_{n \to \infty}(S_n - G_n)$ exists. What is the value of the limit?

*Solution by Chico Problem Group, California State University, Chico.* With the proposer's notation (one suspects he intended to label $S_n$ and $G_n$ in the opposite way) we show

$$\lim_{n \to \infty}(S_n - G_n) = -\frac{5}{24} - \zeta\left(-\frac{1}{2}\right),$$

where $\zeta(s)$ is the Riemann zeta function.

First, rewriting $S_n$ with the aid of Taylor's formula

$$\sqrt{n + \frac{1}{4}} = \sqrt{n} + \frac{1}{8\sqrt{n}} + O(n^{-3/2}),$$

we have

$$S_n = \frac{2}{3}\left(n + \frac{5}{8}\right)\left(n + \frac{1}{4}\right)^{1/2} - \frac{5}{24}$$

$$= \frac{2}{3}n\left(\sqrt{n} + \frac{1}{8\sqrt{n}}\right) + \frac{5}{12}\sqrt{n} - \frac{5}{24} + o(1)$$

$$(1) \qquad = \frac{2}{3}n^{3/2} + \frac{1}{2}n^{1/2} - \frac{5}{24} + o(1).$$

Hence,

$$(2) \qquad \lim_{n\to\infty}(S_n - G_n) = \lim_{n\to\infty}\left(\frac{2}{3}n^{3/2} + \frac{1}{2}n^{1/2} - G_n\right) - \frac{5}{24}$$

provided the latter limit exists.

From the Euler-Maclaurin summation formula we have

$$G_n = \frac{1}{2}(1 + \sqrt{n}) + \int_1^n \sqrt{x}\,dx + \frac{1}{2}\int_1^n \frac{P_1(x)}{x^{1/2}}\,dx,$$

where $P_1(x) = x - [x] - \frac{1}{2}$. Thus

$$(3) \qquad G_n = \frac{2}{3}n^{3/2} + \frac{1}{2}n^{1/2} - \frac{1}{6} + \frac{1}{2}\int_1^n \frac{P_1(x)}{x^{1/2}}\,dx.$$

The periodicity and alternating sign properties of $P_1(x)$ insure that $\int_1^\infty P_1(x)\phi(x)\,dx$ is finite for any positive, continuous function $\phi(x)$ which decreases monotonically to zero as $x \to +\infty$. Thus we can write (3) as

$$G_n = \frac{2}{3}n^{3/2} + \frac{1}{2}n^{1/2} - \frac{1}{6} + \frac{1}{2}\int_1^\infty \frac{P_1(x)}{x^{1/2}}\,dx + o(1),$$

and comparing this with (2) we have

$$(4) \qquad \lim_{n\to\infty}(S_n - G_n) = -\frac{1}{24} - \frac{1}{2}\int_1^\infty \frac{P_1(x)}{x^{1/2}}\,dx$$

with the limit clearly existing. It remains to evaluate this last integral.

By a standard method for continuing the zeta function analytically ($\Sigma_1^\infty n^{-s}$ is written as the Stieltjes integral $\int_1^\infty x^{-s}d(x - [x])$ which is then integrated by parts) we have the formula

$$(5) \qquad \zeta(s) - \frac{1}{s-1} = 1 - s\int_1^\infty \frac{x - [x]}{x^{s-1}}\,dx,$$

where the right-hand side is defined and analytic for $\mathrm{Re}(s) > 0$.

Replacing the numerator in the integral of (5) by $P_1(x) = x - [x] - \frac{1}{2}$ gives the formula

$$(6) \qquad \zeta(s) - \frac{1}{s-1} = \frac{1}{2} - s\int_1^\infty \frac{P_1(x)}{x^{s+1}}\,dx,$$

where now the integral converges and represents an analytic function for $\mathrm{Re}(s) > -1$. Substituting $s = -\frac{1}{2}$ in (6) yields

$$\int_1^\infty \frac{P_1(x)}{x^{1/2}}\,dx = \frac{1}{3} + 2\zeta\left(-\frac{1}{2}\right),$$

which along with (3) gives

$$\lim_{n\to\infty}(S_n - G_n) = -\frac{5}{24} - \zeta\left(-\frac{1}{2}\right).$$

No closed form expressions for $\zeta(-\frac{1}{2})$ are known. A decimal approximation [1] is

$$\zeta\left(-\frac{1}{2}\right) \approx -0.20788\ 62249\ 77354\ 56602,$$

so that

$$\lim_{n \to \infty} (S_n - G_n) \approx -.00044\ 71083\ 55978\ 76732.$$

### Reference

1. J. W. Wrench, Jr., private communication, cf. Math Comp., V. 22 (1968) 687–688.

Also solved by P. S. Bruckman, G. H. Gonnet (Canada), T. K. Louton and C. C. Rousseau, O. P. Lossers (The Netherlands), W. A. Newcomb, A. Stenger, S. K. Venkatesan (India), P. Y. Wu (Taiwan), and the proposer.

### A Triangle Inequality

E 2958 [1982, 498]. *Proposed by M. S. Klamkin, University of Alberta.*

Let $x, y, z$ be positive, and let $A, B, C$ be angles of a triangle. Prove that $x^2 + y^2 + z^2 \geqslant 2yz\sin(A - \pi/6) + 2zx\sin(B - \pi/6) + 2xy\sin(C - \pi/6)$.

*Solution by C. S. Karuppan Chetty, Regional Engineering College, Tiruchirapalli, India.*

*Solution* I. Let

$$f(x, y, z) = x^2 + y^2 + z^2 - 2yz\sin\alpha - 2zx\sin\beta - 2xy\sin\gamma,$$

where $\alpha = A - \pi/6$, $\beta = B - \pi/6$ and $\gamma = C - \pi/6$. Since

$$\sin\alpha = \cos\beta\cos\gamma - \sin\beta\sin\gamma,$$

we can express $f(x, y, z)$ as

$$(x - y\sin\gamma - z\sin\beta)^2 + (y\cos\gamma - z\cos\beta)^2$$

and the result follows.

In fact, the given inequality holds for any real $x, y, z$ and for all $A, B, C$ with $A + B + C = \pi$.

*Solution* II. Defining $f(x, y, z)$ as above, the matrix of the quadratic form $f$ is

$$U = \begin{bmatrix} 1 & -\sin\gamma & -\sin\beta \\ -\sin\gamma & 1 & -\sin\alpha \\ -\sin\beta & -\sin\alpha & 1 \end{bmatrix}.$$

It is easily seen that $\det(U) = 0$, and hence the leading principal minors of $U$ are 1, $\cos^2\gamma$ and 0. Therefore, $f$ is positive semidefinite and the result follows.

Also solved by K. L. Bernstein, K. V. Bhagwat (India), R. Breusch, Z. A. L. Geöcze (Brazil), R. Heller, W. Janous (Austria), H. Kestelman (England), L. Kuipers (Switzerland), M. J. Lempel, O. P. Lossers (The Netherlands), A. Marguina (Spain), V. D. Mascioni (Switzerland), W. A. Newcomb, T. Riessinger (West Germany), K. Rogers, O. G. Ruehr, I. A. Sakmar (Canada), K. L. Stellmacher, G. A. Tsintsifas (Greece), M. Vowe (Switzerland), and the proposer. Partial solutions by M. Bencze (Romania) and E. Braune (West Germany).

### Inradii-circumradii Products for Tetrahedra

E 2962 [1982, 593]. *Proposed by M. S. Klamkin, University of Alberta, Canada.*

It is known that if the circumradii $R$ of the four faces of a tetrahedron are congruent, then the four faces of the tetrahedron are mutually congruent (i.e., the tetrahedron is isosceles) [**1**]. It is also known that if the inradii $r$ of the four faces of a tetrahedron are congruent, then the tetrahedron need not be isosceles [**2**]. Show that if $Rr$ is the same for each face of a tetrahedron, the tetrahedron is isosceles.

### References

1. Crux Mathematicorum, 6 (1980) 219.
2. Crux Mathematicorum, 4 (1978) 263.

*Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* Let $ABCD$ be a tetrahedron. We put $\overline{BC} = a$, $\overline{CA} = b$, $\overline{AB} = c$, $\overline{DA} = a_1$, $\overline{DB} = b_1$, $\overline{DC} = c_1$. The fact that the product of the circumradius and the inradius is the same for each face is then expressed by

$$\frac{abc}{a+b+c} = \frac{ab_1c_1}{a+b_1+c_1} = \frac{a_1bc_1}{a_1+b+c_1} = \frac{a_1b_1c}{a_1+b_1+c} = 2Rr.$$

The first equality is equivalent to

$$bb_1(c-c_1) + cc_1(b-b_1) = a(b_1c_1 - bc),$$

or

$$c_1(a+c)(b-b_1) + b(b_1+a)(c-c_1) = 0. \tag{1}$$

We have two more equations obtained from (1) by cyclic permutation, namely

$$c(c_1+b)(a-a_1) + a_1(b+a)(c-c_1) = 0, \tag{2}$$

$$b_1(b+c)(a-a_1) + a(a_1+c)(b-b_1) = 0. \tag{3}$$

Considering (1), (2) and (3) as a system of linear equations for the unknowns $a - a_1, b - b_1, c - c_1$ we observe that the determinant of the system has the form

$$\begin{vmatrix} 0 & p & q \\ r & 0 & s \\ t & u & 0 \end{vmatrix} = pst + qru$$

and does not vanish because all its elements are obviously positive. Hence $a = a_1$, $b = b_1$, $c = c_1$, which proves the statement.

Also solved by L. Bankoff, J. Heuver (Canada), M. R. Railkar (India), G. Tsintsifas (Greece), and the proposer.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by August 31, 1985. The solver's full post-office address should be on each sheet.*

6493. *Proposed by M. L. Glasser, Clarkson College.*

Show that for $a_j > 0$

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_k \frac{\sin a_1x_1}{x_1} \cdots \frac{\sin a_kx_k}{x_k} \frac{\sin a_0(x_1 + \cdots + x_k)}{x_1 + \cdots + x_k} = \pi^k \min\{a_0, \ldots, a_k\}$$

(see 5314 [1965, 759]).

6494. *Proposed by A. Wilansky, Lehigh University.*

Let $X$ be a commutative Banach algebra with 1 and $M, N$ maximal ideals. Let $S$ be the linear span of $\{mn\colon m \in M,\ n \in N\}$. Show that the codimension of $S$ must be finite if $M \neq N$. (This is not necessarily true if $M = N$.)

6495. *Proposed by Robert E. Shafer, Berkeley, California.*

For $x > 1$, define

$$\psi(x) = \lim_{N\to\infty}\left\{\log(x+N) - \sum_{n=0}^{N} \frac{1}{x+n}\right\},$$

$$P(x) = \lim_{N\to\infty}\left\{\log\log(x+N) - \sum_{n=0}^{N}\frac{1}{(x+n)\log(x+n)}\right\}.$$

Prove that $P(x) > \log\psi(x)$ for $x$ sufficiently large.

## SOLUTIONS OF ADVANCED PROBLEMS

### A Right-Angle Preserving Mapping

6436 [1983, 485]. *Proposed by the late H. Kestelman, University College, London.*

Let $V$ be a real vector space normed by a scalar product. Let $f: V \to V$ be such that whenever $a, b, c$ are vertices of a triangle right-angled at $a$, then $f(a), f(b), f(c)$ are vertices of a triangle right-angled at $f(a)$. Show that $f$ is a continuous collineation.

*Solution by Alain Tissier, Montfermeil, France.* We assume that the dimension of $V$ is at least 2.

LEMMA 1. *Suppose*

(1) $\quad a_1, a_2, a_3, a_4$ *are the vertices of a square with centre* $b$.

*Then* $f(a_1)$, $f(a_2)$, $f(a_3)$, $f(a_4)$ *are the vertices of a square with centre* $f(b)$.

*Proof.* Clearly (1) implies

(2) $\quad (a_i - b)\cdot(a_{i+1} - b) = 0$ and $(a_i - a_{i+1})\cdot(a_{i+1} - a_{i+2}) = 0$ for $1 \leqslant i \leqslant 4$,

where $a_{i+4} = a_i$. Conversely it follows from (2), by the Pythagorean theorem, that, for $1 \leqslant i \leqslant 4$,

$$\|a_i - b\|^2 + \|a_{i+1} - b\|^2 = \|a_i - a_{i+1}\|^2$$

and

$$\|a_i - a_{i+1}\|^2 + \|a_{i+1} - a_{i+2}\|^2 = \|a_i - a_{i+2}\|^2,$$

whence, by an easy computation,

$$\|a_1 - a_3\| = \|a_2 - a_4\| =: \delta,$$

and then

$$\|a_{i+1} - a_i\| = \delta/\sqrt{2}, \|a_1 - b\| = \|a_3 - b\| =: \alpha,$$

and

$$\|a_2 - b\| = \|a_4 - b\| =: \beta.$$

But, by the triangle inequality, $\delta \leqslant 2\alpha$ and $\delta \leqslant 2\beta$, so that $\alpha = \beta = \delta/2$. Thus (2) is equivalent to (1). Next, if follows from (2) and the right-angle preserving property of $f$ that (2) holds with $f(a_i), f(b)$ in place of $a_i, b$ respectively. This completes the proof of Lemma 1.

Let $\Delta$ be the set of numbers in $[0,1]$ of the form $m/2^n$, where $m, n$ are nonnegative integers.

LEMMA 2. *If* $a, b \in V$, $\lambda \in \Delta$, *then* $f((1-\lambda)a + \lambda b) = (1-\lambda)f(a) + \lambda f(b)$.

*Proof.* Given $a, b \in V$, there are points $c, d \in V$ such that $a, c, b, d$ are the vertices of a square. Hence, by Lemma 1,

$$f\left(\frac{a+b}{2}\right) = \frac{1}{2}(f(a) + f(b)).$$

The proof can now be easily completed by induction.

LEMMA 3. *For each* $a \in V$ *there is a mapping* $h_a$: $[0,\infty) \to [0,\infty)$ *such that* $\|x-a\| = \rho$ *implies* $\|f(x)-f(a)\| = h_a(\rho)$.

*Proof*. Given $a \in V$, $\rho \geqslant 0$, there are points $b, c \in V$ such that

$$\|b-a\| = \|c-a\| = \rho \quad \text{and} \quad a = (b+c)/2.$$

Then $\|x-a\| = \rho$ is equivalent to $(x-b)\cdot(x-c) = 0$ which implies

$$(f(x)-f(b))\cdot(f(x)-f(c)) = 0.$$

Since $f(a) = (f(b)+f(c))/2$, it follows that $\|x-a\| = \rho$ implies

$$\|f(x)-f(a)\| = \|f(a)-f(b)\|$$

which depends on $a$ and $\rho$ but not on $x$.

LEMMA 4. *If* $a \in V$ *and* $0 \leqslant r < \rho$, *then* $h_a(r) < h_a(\rho)$.

*Proof*. There are points $b, c \in V$ such that

$$\|c-a\| = r < \|b-a\| = \rho \quad \text{and} \quad (c-a)\cdot(c-b) = 0.$$

Hence

$$(f(c)-f(a))\cdot(f(c)-f(b)) = 0$$

and so, by the Pythagorean theorem,

$$h_a(\rho)^2 = \|f(a)-f(b)\|^2 = \|f(c)-f(a)\|^2 + \|f(c)-f(b)\|^2$$
$$> \|f(c)-f(a)\|^2 = h_a(r)^2.$$

To prove $f$ is continuous, it suffices, in view of Lemmas 3 and 4, to show that $\lim_{n\to\infty} h_a(2^{-n}\rho) = 0$ for every $a \in V$, $\rho > 0$. This is a consequence of Lemma 2 which yields, for $\|a-b\| = \rho$ and $c = (1-2^{-n})a + 2^{-n}b$, that

$$h_a(2^{-n}\rho) = \|f(c)-f(a)\| = 2^{-n}\|f(b)-f(a)\| = 2^{-n}h_a(\rho).$$

Since $\bar{\Delta} = [0,1]$, it follows from Lemma 2 that

$$f((1-\lambda)a+\lambda b) = (1-\lambda)f(a)+\lambda f(b)$$

for every $\lambda \in [0,1]$ and hence, by a simple argument, that the identity holds for every real $\lambda$. This shows that $f$ is a collineation.

Finally, it is easy to deduce from the above that $\|f(a)-f(x)\| = k_a\|a-x\|$ where $k_a$ is independent of $x$, i.e., $f$ is a similarity.

Also solved by Aage Bondesen (Denmark) and the proposer, and partially solved by L. E. Mattics.

### An Exponential Sum

6442 [1983, 648]. *Proposed by Vladimir Naroditsky, San Jose State University.*

Show that if

$$\lim_{t\to\infty} \sum_{k=1}^{N} a_k \exp(i\lambda_k t)$$

exists, then either all $\lambda$'s are 0 or all $a$'s are 0.

*Solution by Piotr Biler, Institute of Mathematics, University of Wrocław, Poland.* The proposition as stated is evidently incorrect. For example, the sum is constant if either $a_k\lambda_k = 0$ for $k = 1, 2, \ldots, N$ or all the $\lambda$'s are equal and $a_1 + a_2 + \cdots + a_N = 0$. Suppose that the $\lambda$'s are all

different nonzero real numbers. Denote the sum by $S(t)$ and its limit by $S$. Since $T^{-1}\int_0^T S(t)\,dt \to 0$ as $T \to \infty$, it follows that $S = 0$. Hence, for $k = 1, 2, \ldots, N$, $S(t)\exp(-i\lambda_k t) \to 0$ as $t \to \infty$ and so

$$0 = \lim_{T\to\infty} \frac{1}{T}\int_0^T S(t)\exp(-i\lambda_k t)\,dt = a_k.$$

Also solved by Preben Alsholm, Bennett Eisenberg, J. A. Goldstein, Michael B. Gregory, J. P. Henniger (Canada), N. J. Lord (United Kingdom), Jacques Moisan (France), Jean-Marie Monier (France), William A. Newcomb, M. Pachter (South Africa), Abraham Smuckler (Israel), Allen Stenger, Alain Tissier (France), Wolfgang Walter (West Germany), and the proposer.

### A Probability Inequality

6443 [1983, 648]. *Proposed by C. Thorn and B. Tomaszewski, University of Wisconsin.*

Let $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_{2n+1} \geqslant 0$ be a decreasing sequence of positive real numbers and let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{2n+1}$ be a Bernoulli sequence of independent random variables, i.e., $P(\varepsilon_i = 1) = P(\varepsilon_i = -1) = \frac{1}{2}$. Prove that

$$P\left(\left|\sum_{i=1}^{2n+1}\varepsilon_i a_i\right| < a_1\right) + \frac{1}{2}P\left(\left|\sum_{i=1}^{2n+1}\varepsilon_i a_i\right| = a_1\right) \geqslant \frac{1}{2^{2n+1}}\binom{2n+1}{n}.$$

*Solution by Lajos Takács, Case Western Reserve University, Cleveland, Ohio.* Let $\{b_n\}$ be a nondecreasing sequence of positive real numbers, let $s_0 = 0$ and, for $n = 1, 2, \ldots$, let

$$s_n = b_1\varepsilon_1 + b_2\varepsilon_2 + \cdots + b_n\varepsilon_n, \qquad t_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n.$$

Let

$$Q_n(x) = P(s_n > x) + \frac{1}{2}P(s_n = x).$$

Then the required inequality can be expressed as

$$2Q_n(b_n) \leqslant 1 - \frac{1}{2^n}\binom{n}{[n/2]}$$

when $n$ is an odd integer. We shall in fact show the inequality to hold for every positive integer $n$. Evidently

$$2Q_n(x) = Q_{n-1}(x - b_n) + Q_{n-1}(x + b_n) \quad \text{for } n = 1, 2, \ldots.$$

Since $Q_n(x) \geqslant Q_n(y)$ when $x \leqslant y$, we have

$$2Q_n(jb_n) \leqslant Q_{n-1}((j-1)b_{n-1}) + Q_{n-1}((j+1)b_{n-1}) \quad \text{for } n, j = 1, 2 \ldots.$$

Hence $Q_n(jb_n) \leqslant f_n(j)$ for $n, j = 0, 1, \ldots,$ where

$$2f_n(j) = f_{n-1}(j-1) + f_{n-1}(j+1),$$

and $f_n(0) = 1/2$ for $n = 0, 1, \ldots, f_0(j) = 0$ for $j = 1, 2, \ldots$. Since $f_n(j) = Q_n(j)$ in the case where $b_1 = b_2 = \cdots = b_n = 1$, we obtain that

$$\begin{aligned} 2f_n(1) &= 2P(t_n > 1) + P(t_n = 1) = 1 - P(t_n = 1) \\ &= 1 - \frac{1}{2^n}\binom{n}{[n/2]} \text{ for } n = 1, 2, \ldots, \end{aligned}$$

and the required inequality follows.

Also solved by the proposers.

**On Landen and Pfaff Transformations**

6445 [1983, 709]. *Proposed by Richard Askey, University of Wisconsin.*

Show that Pfaff's transformation of the hypergeometric function

$$ {}_2F_1\left(\begin{matrix}a,b\\c\end{matrix};x\right)=(1-x)^{-a}{}_2F_1\left(\begin{matrix}a,c-b\\c\end{matrix};\frac{x}{x-1}\right)$$

implies Landen's transformation of the dilogarithm

$$\mathrm{Li}_2x:=\sum_{n=1}^{\infty}\frac{x^n}{n^2}=-\frac{1}{2}[\log(1-x)]^2-\mathrm{Li}_2(x/(x-1)).$$

*Solution by Mourad E. H. Ismail, Arizona State University, Tempe, Arizona.* It is easy to see that, for $-1\leqslant x<1$,

$$\log(1-x)=-x\,{}_2F_1\left(\begin{matrix}1,1\\2\end{matrix};x\right)\text{ and }\mathrm{Li}_2x=\int_0^x\sum_0^{\infty}\frac{t^n}{n+1}dt=\int_0^x{}_2F_1\left(\begin{matrix}1,1\\2\end{matrix};t\right)dt.$$

The Pfaff transformation and the above integral representation give, for $-1\leqslant x\leqslant 1/2$,

$$\begin{aligned}\mathrm{Li}_2(x/(x-1))&=\int_0^{\frac{x}{x-1}}{}_2F_1\left(\begin{matrix}1,1\\2\end{matrix};\frac{t}{t-1}\right)\frac{dt}{1-t}=-\int_0^x{}_2F_1\left(\begin{matrix}1,1\\2\end{matrix};u\right)\frac{du}{1-u}\\&=\int_0^x\frac{\log(1-u)}{u(1-u)}du=\int_0^x\left(\frac{1}{u}+\frac{1}{1-u}\right)\log(1-u)\,du\\&=-\mathrm{Li}_2x-\frac{1}{2}[\log(1-x)]^2.\end{aligned}$$

Also solved by Paul S. Bruckman, Henry E. Fettis, C. Georghiou (Greece), Ira Gessel, M. L. Glasser, L. Kuipers (Switzerland), William A. Newcomb, Robert E. Shafer, and the proposer.

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Differential Topology: An Introduction.* By David B. Gauld. Marcel Dekker, New York, 1982. v + 241 pp. $29.75.

THEODORE SHIFRIN
*Department of Mathematics, University of Georgia, Athens, GA 30602*

The basic objects of study in differential topology are the *differentiable* (or *smooth*) *manifolds* and the differentiable functions on them. The prototype is the hypersurface $X=\{f(x,y,z,)=0\}$ in $\mathbb{R}^3$, where $f(x,y,z)$ is differentiable and $\vec{\nabla}f\neq 0$ on $X$, i.e., $X$ contains no critical point of $f$. Surfaces, whether topological, polyhedral, or smooth, have interested many mathematicians, among them Descartes, Euler, Riemann and Poincaré. Smooth manifolds appear in all walks of life—as state and phase spaces in physics and chemistry, as indifference sets in economics, as orbit spaces in dynamical systems as space-time in general relativity, and so forth (cf. [**1**]). Problems in the calculus of variations, numerical analysis and applied mathematics are often phrased in terms of functions on manifolds (both finite- and infinite-dimensional).

Roughly speaking, we may define a smooth $n$-dimensional manifold as follows. Topologically,

it is a locally Euclidean metric space: each point has a neighborhood which is homeomorphic to $\mathbb{R}^n$. Analytically, it admits calculus: one knows which functions ought to be called differentiable, and one has a well-defined tangent space at each point, as for the surfaces in multivariable calculus. It turns out that every manifold $X$ may be viewed as sitting in some (large) Euclidean space $\mathbb{R}^N$; then it is not hard to prove that locally $X$ is the zero locus of $N$-$n$ independent functions (i.e., their gradient vectors are linearly independent along $X$).

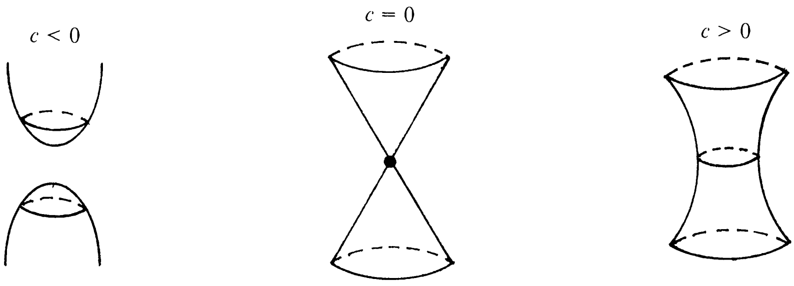



FIG. 1

Consider the family of surfaces $\{f(x,y,z) = x^2 + y^2 - z^2 = c\}$. (See Fig 1.) At $c = 0$ the topological type of the surfaces changes dramatically with a small change in $c$, but for $c \neq 0$ the surfaces exhibit stability. What is so special about the case $c = 0$? Let $f\colon \mathbb{R}^n \to \mathbb{R}^k$ be differentiable; we say $y \in \mathbb{R}^k$ is a *regular value* of $f$ if rank $(df_x) = k$ for every $x \in f^{-1}(y)$; otherwise, $y$ is a *critical value* of $f$. The salient facts are the following: (1) (Implicit function theorem) If $y$ is a regular value of $f$, then $f^{-1}(y)$ is a smooth submanifold of $\mathbb{R}^n$. (2) (Sard's theorem) The set of regular values of $f$ is open and dense in $\mathbb{R}^k$. Here we see already the importance of two fundamental concepts in differential topology, as in mathematics generally: *stability* (openness) and *genericity* (denseness). We can now propose two routes to follow. The first is to study geometric properties arising from the generic or stable situation. The second is to infer information from the singularities themselves.

In the first direction, one of the most elementary concepts is that of the *degree* of a mapping between manifolds of the same dimension—roughly, the number of times (counting orientations) that the domain covers the target under the mapping. The notion was first introduced early in the century by Brouwer from a combinatorial viewpoint, but it is a very natural setting for the differentiable viewpoint. Given a smooth map $f\colon X \to Y$, with $X$ compact and $Y$ connected, choose a regular value $y$ of $f$ and count the (finite) number of points in $f^{-1}(y)$ with signs. One concludes from the inverse function theorem and the topological assumptions we've made that this number is independent of the choice of $y$ and is, indeed, invariant under perturbations of $f$ (after all, how can an integer-valued function vary continuously?). As an elementary application of this, we mention the famous hairy ball theorem: any (tangent) vector field on the two-sphere $S^2$ must vanish somewhere. For if not, we could deform the identity map to the antipodal map by pushing in the direction of the vector field, and so the two mappings would have to have the same degree. As it happens, the identity map has degree $+1$ and the antipodal map, $-1$.

In a similar vein, there is another famous elementary topological result. Call a map $f\colon S^2 \to S^2$ *even* if $f(-x) = f(x)$, and *odd* if $f(-x) = -f(x)$. It is not hard to see that an even map has even degree; what is not so obvious is that an odd map must have odd degree. From this one may deduce the fact that if $g\colon S^2 \to \mathbb{R}^2$ is continuous, there is an $x \in S^2$ with $g(x) = g(-x)$; more colloquially, at any given time, there is a pair of antipodal points on the earth where the temperature and relative humidity are precisely equal.

Once one has the idea of counting points, it is natural to broach the subject of intersection theory which plays a central role in topology, geometry, and algebraic geometry (and, indeed, as

index theory, in analysis). Given a manifold $X$ and compact manifolds $Y, Z \subset X$, then it is a consequence of Sard's theorem above that, after an arbitrarily small deformation of $Z$ (say) in $X$, we may insure that $Y$ and $Z$ meet *transversely*, or *in general position* (i.e., so that the sum of their tangent spaces is the tangent space of the ambient space); it follows once again from the implicit function theorem that then $Y \cap Z$ is a smooth submanifold of $X$. (See Fig. 2.) As a simple case of this, most mathematicians have at one time used the general position lemma that the slice of a surface in $\mathbb{R}^3$ by the general plane is a smooth curve (in algebraic geometry, this comes under the guise of Bertini's theorem and has far-reaching consequences).

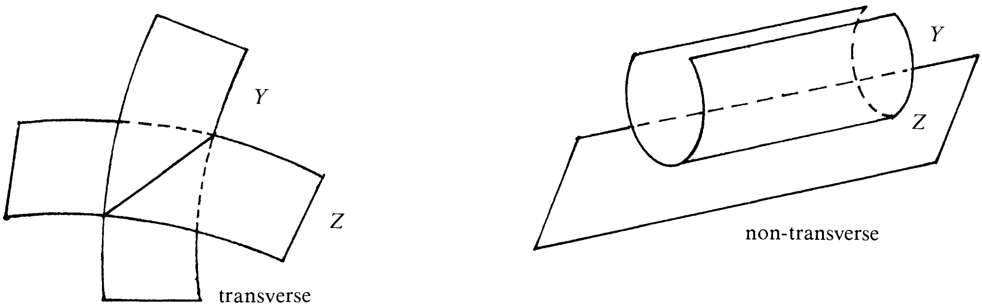



FIG. 2

One particular case of intersection theory is that of complementary dimension, and when $X, Y, Z$, are all oriented, we may speak of the intersection number of $Y$ and $Z$ (in $X$). Such numbers turn out to be indicative of the global nature of $X$. For example, the classical Euler characteristic $\chi(X)$ ($= V - E + F$ for a polyhedral surface) appears in just such a fashion; the link here is the beautiful result (due to Poincaré and Hopf) that this number is equal to the number of zeroes of a vector field on the manifold, counted appropriately. And one of the prettiest and most inspiring theorems of this century is Chern's generalization of the classical Gauss-Bonnet theorem, which equates the Euler characteristic of an even-dimensional compact oriented Riemannian manifold to a curvature integral (cf. [**4**], [**8**], [**12**]).

The goal of much mathematical work is, after all, the classification problem. The thrust of topology is to develop a list of properties (e.g., compactness, connectedness, separability) which distinguish among topological spaces. Algebraic topology translates this general problem into algebra—homology, homotopy, cohomology, bordism—and intersection theory is one geometric version of this (indeed, it was the classical approach of Poincaré and Lefschetz). In light of Hopf's result, we see that the Euler characteristic of a connected manifold $X$ gives the *obstruction* to

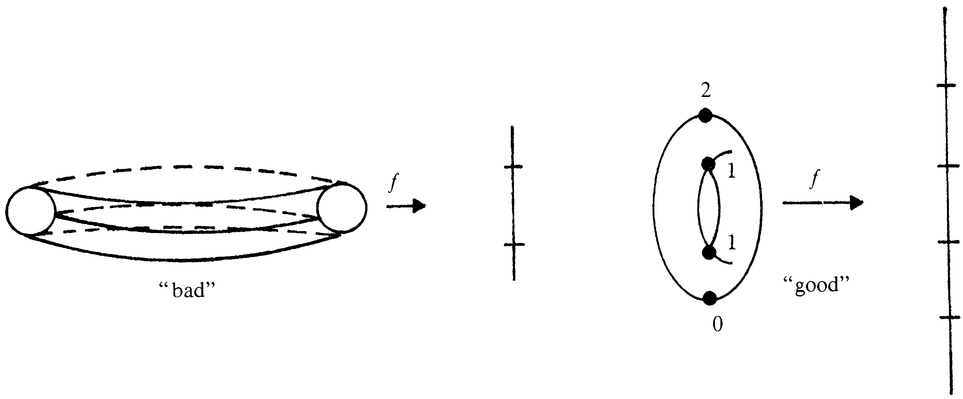



FIG. 3

finding a vector field on $X$ which is nowhere zero; it is a beautiful fact that the converse holds as well: if $\chi(X) = 0$, then there exists a nowhere-zero vector field on $X$ (cf. [6] or [7]).

Let us turn now to our second point of view; we wish to glean information about a manifold $X$ from the singularities of a nice function defined on it. This is the domain of Morse theory. Let $X \subset \mathbb{R}^N$ be a compact $n$-manifold, and consider the height function $f(x) = x_N$ on $X$ (see Fig. 3). By rotating the coordinate system slightly if necessary, we may insure that $f$ is a function with only nondegenerate critical points. That is, in any local coordinate system $(z_1, \ldots, z_n)$ on $X$ centered at a critical point $p$, $\partial f/\partial z_i(0) = 0$ and the hessian matrix

$$H = \left[ \frac{\partial^2 f}{\partial z_i \partial z_j}(0) \right]$$

is nonsingular. Such a function $f$ is called a *Morse function* on $X$. Note, by the way, that the critical points of $f$ are precisely the points where the hyperplanes $x_N$ = constant fail to meet $X$ transversely. Now, the number of negative eigenvalues of $H$ is well-defined and is called the *index* of the critical point $p$. In Fig. 3 we have labelled the indices of the Morse function on the right.

Let us begin by observing that as we pass through critical values, the topology of the preimage sets $f^{-1}(y)$ changes dramatically; this is the same phenomenon with which we began in Fig. 1.

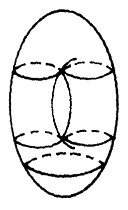


FIG. 4

(See Fig. 4.) Given our Morse function on $X$, we would like a prescription for concocting $X$ out of "building blocks." Here we adopt the simplest viewpoint. Let $D^k = \{x \in \mathbb{R}^k\colon \|x\| \leq 1\}$; we call this a $k$-cell; its boundary is the standard $(k-1)$-sphere, $\partial D^k = S^{k-1}$. In the purely topological category, ignoring differentiability, we can speak of a cell complex, a topological space built inductively by attaching $k$-cells $D^k$ to $X_{k-1}$, the collection of all cells of dimension $k-1$, by a continuous map $\phi\colon \partial D^k \to X_{k-1}$. For example, we may obtain the $n$-sphere $S^n$ by attaching an $n$-cell to a 0-cell, as pictured in Fig. 5:

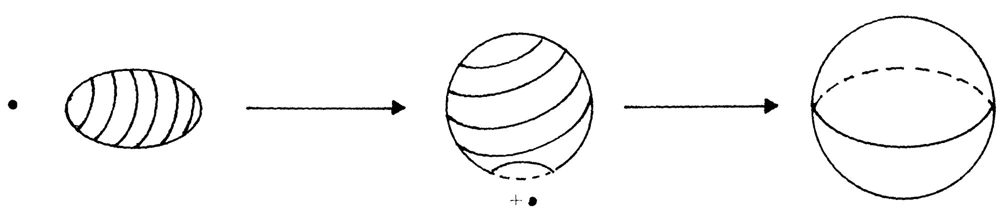


FIG. 5

The fundamental theorem here, linking algebraic and differential topology, states that any compact manifold $X$ may be built out of cells, attaching one $k$-cell for each critical point of index $k$ (for a given Morse function on $X$). Note that this gives many different cell structures for the same underlying space. For example, in Fig. 6 on p. 298,

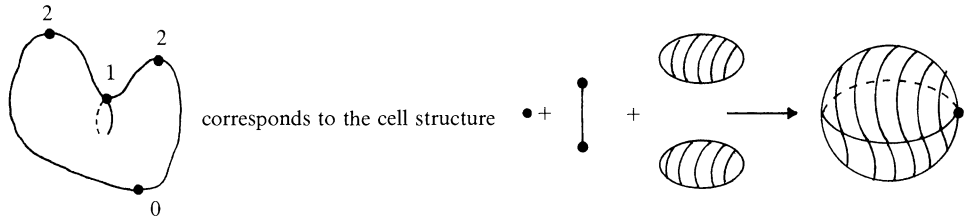



FIG. 6

for the sphere. Similarly, the standard torus with the Morse function in Fig. 3 is built from the standard picture shown in Fig. 7:

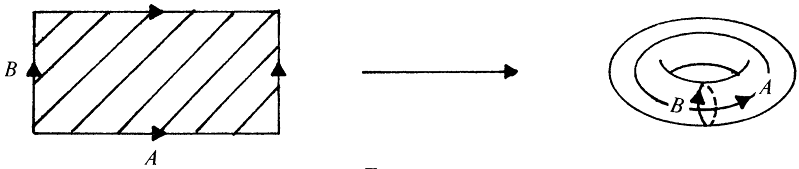



FIG. 7

The theory is very pretty and very geometric; nevertheless, to prove such a theorem, one must use the integral curves of a differential equation derived from the Morse function.

Surgery theory is based on the following innocent observation. Suppose we have a copy of $S^k \times D^{n-k}$ sitting in an $n$-manifold $X$; then, keeping its boundary the same, we can replace its interior by $D^{k+1} \times S^{n-k-1}$, since, after all,

$$\partial(D^{k+1} \times S^{n-k-1}) = S^k \times S^{n-k-1} = \partial(S^k \times D^{n-k}).$$

This process, for example, turns the 2-sphere into the standard torus, as shown in Fig. 8:

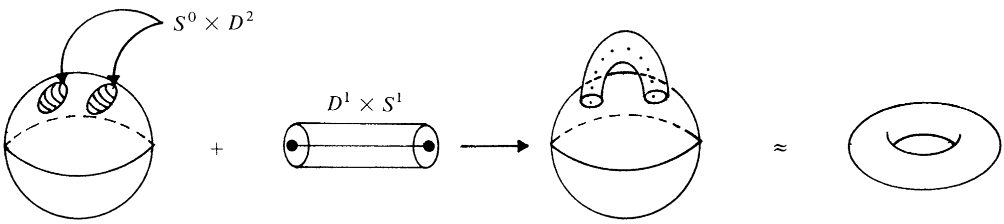



FIG. 8

This seems an obvious, yet puzzling construction; for normally in mathematics, one performs modifications which preserve the invariants one is trying to study. But surgery blatantly changes the topology of the space. That ultimately is the point! We referred earlier to obstructions—surgery may be used to remove them, or at least to make them more manageable. It was to this end that Milnor, Browder, and Novikov, among others, invented the theory. For example, the high-dimensional Poincaré conjecture states that if $X$ is a compact $n$-manifold without boundary, $n \geq 5$, with the homotopy type of a sphere, then $X$ is homeomorphic to $S^n$. Indeed, in dimension five (and six), $X$ is diffeomorphic to $S^5$ (resp., $S^6$). The main point here is to prove that $X$ is the boundary of a 6-manifold $W$ that is contractible (i.e., homotopically a point); first one finds any old $W$ (more or less), and then, using surgery techniques, one "kills" all its homotopy groups. So we have altered $W$, but produced a new manifold with boundary $X$ which satisfies the prerequisites. Then the result follows from Smale's $h$-cobordism theorem (cf. [**10**], [**13**]).

We have mentioned only a few of the problems with which differential topology deals, largely

because these are the notions which the undergraduate texts currently available treat. We've omitted Smale's handlebody theory, Thom's work on cobordism (cf. [**13**]), all the ramifications of singularity theory, which has become a huge industry recently, intertwining algebraic geometry and differential topology (cf. [**5**], [**2**], [**3**]); we've also omitted immersion theory, characteristic class theory (cf. [**12**]), and of course the celebrated work on 3- and 4-manifolds recently by Kirby, Thurston and their schools.

David Gauld's book allegedly presumes no knowledge of point-set topology or advanced calculus, starting with a cursory treatment of both and then proceeding to the definitions of manifolds and Morse functions. I find it hard to believe that the undergraduate initiate can successfully handle the technicalities involved in the latter half of the book unless he is already well-grounded in both prerequisites. Gauld's treatment of elementary surgery theory is, however, readable and almost too full of detail; perhaps it will make a good reference in that regard for a graduate student learning the theory. The book finishes with a treatment of the classification of compact, oriented surfaces.

I am puzzled by Gauld's bibliographical notes. He says that Wallace's book *Differential Topology: First Steps* ([**14**]) "although different in aim and scope from this book, [is] at a similar level and covers more or less the same topics." Indeed, having read both books with reasonable care, I can't imagine how the aim and scope are different. Wallace covers more material in a terser fashion, and gives the reader a greater feeling for the impact of surgery theory. It seems to me that Gauld is more willing to get bogged down in the technical details.

Surgery theory is unquestionably a beautiful and important part of modern mathematics; nevertheless, I find the subject and this book less appropriate for an undergraduate course than transversality theory and its applications. Admittedly, I am biased, having taught an undergraduate course based on [**6**] and [**11**] three times. These days, more of our students in pure mathematics courses come from computer science, physics, mathematical economics, and so forth. I believe this makes the course I prefer to teach even more appropriate, because it integrates linear algebra, advanced calculus, and point-set topology very strongly, and builds foundational strength by applying the transversality theory over and over again, obtaining a number of deep topological results meanwhile. It is true that the structure theory of manifolds, i.e., the geometry of Morse theory, is all but missing; but on the other hand, to actually do surgery theory takes plenty of transversality arguments (which Wallace mentions briefly) and much algebraic topology. But, indeed, I would like to campaign for such an undergraduate differential topology course in our curriculum. Unfortunately, mathematics courses at the undergraduate level tend to be far too independent from one another; this particular subject affords us the opportunity to demonstrate to our students the beauty in the unity of mathematics.

**References**


**1.** F. Browder (ed.), The Mathematical Heritage of Henri Poincaré, Proc. Symp. Pure Math, vol. 39 (1983).

**2.** J. Callahan, Singularities and plane maps, this MONTHLY, 81 (1974) 211–240.

**3.** ______, Singularities and plane maps, II, this MONTHLY, 84 (1977) 765–803.

**4.** S.-S. Chern, A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds, Ann. of Math., 45 (1944) 747–752.

**5.** M. Golubitsky and V. Guillemin, Stable Mappings and Their Singularities, Springer-Verlag, 1973.

**6.** V. Guillemin and A. Pollack, Differential Topology, Prentice-Hall, 1974.

**7.** M. Hirsch, Differential Topology, Springer-Verlag, 1976.

**8.** R. Langevin and T. Shifrin, Polar varieties and integral geometry, Amer. J. Math., 104 (1982) 553–605.

**9.** J. Milnor, Morse Theory, Princeton University Press, 1963.

**10.** ______, Lectures on the $h$-cobordism theorem, Princeton University Press, 1965.

**11.** ______, Topology from the Differentiable Viewpoint, University Press of Virginia, 1965.

**12.** J. Milnor and J. Stasheff, Characteristic Classes, Princeton University Press, 1974.

**13.** S. Smale, A Survey of some recent developments in differential Topology, Bull. Amer. Math. Soc., 69 (1963) 131–145.

**14.** A. Wallace, Differential Topology: First Steps, Benjamin, 1968.

*Examples of Commutative Rings*. By Harry C. Hutchins. Polygonal, Passaic, New Jersey, 1981. vi + 167 pp.

CRAIG HUNEKE
*Department of Mathematics, Purdue University, West Lafayette, IN 47907*

Commutative algebra, as a separate field of mathematics, is comparatively new. Yet already it has become one of the basic structures of mathematics. It arose to complement a trinity of nineteenth century subjects: algebraic number theory, invariant theory, and algebraic geometry.

Like most fields still in a state of development, insight is searched for (and hopefully found!) by a long and careful study of many examples. The book *Examples of Commutative Rings* by Harry C. Hutchins thus provides a useful addition to the basic literature of commutative algebra. It is a good reference to complement a text for a student learning the subject. However, many of the deeper and more revealing examples which would be of interest to researchers are not contained in the book. For instance, important examples which deal with the behavior of rings under completions or group actions are not present, and very little material is devoted to the important topics of resolutions and Hilbert functions.

Much of the present shape of the subject was dictated by problems in number theory and invariant theory. The realization that not all rings of algebraic numbers have the coveted property of unique factorization into prime algebraic numbers led researchers to seek other forms of unique factorization. Kummer successfully completed this program with his ideal numbers. In our later setting, his result states that every ideal in a ring of algebraic integers is uniquely (up to order) a product (or intersection) of prime ideals.

Somewhat later (1890), Hilbert proved in a monumental paper that every ideal in a polynomial algebra is finitely generated, thus obviating to some extent the second main problem of invariant theory. At the same time, Hilbert proved his syzygy theorem, which became largely forgotten, and arose again with the advent of homological algebra, of which more will be said. (A syzygy of a finite collection of elements $f_1, \dots, f_n$ in a ring $R$ is an $n$-tuple $(g_1, \dots, g_n)$ in $R^n$ such that $\Sigma_{i=1}^n g_i f_i = 0$.)

A new challenge was seen: try to extend the factorization of ideals due to Kummer to polynomial rings, the natural setting of both algebraic geometry and invariant theory. This was accomplished in 1905 by the world chess champion, Emmanuel Lasker. He found that every ideal in a polynomial ring is the intersection of primary ideals. (An ideal $I$ is primary if whenever $a \cdot b \in I$, either $a \in I$ or $b^n \in I$ for some $n$.) The result is the best one can do in general. For instance the ideal in $\mathbb{C}[X, Y]$ generated by $X^2$ and $Y^2$ is neither an intersection nor product of prime ideals.

A far-sighted English mathematician, F. S. Macaulay, delved deeper into this decomposition. He was interested in when such a decomposition was unique and "unmixed"—which roughly means all the primary components should have the same "size." Rings where the primary decomposition behaves as well as possible are now called Cohen-Macaulay rings. Cohen-Macaulay rings lie at the foundations of commutative algebra today. In 1974, M. Hochster and J. L. Roberts showed that the ring of invariants of a linearly reductive algebraic group acting on a polynomial ring linearly is Cohen-Macaulay, which with other concurrent results signaled a reinvestigation of invariant theory. A typical example is as follows: let $U$ be an $r \times t$ matrix of variables and $V$ be a $t \times s$ matrix of variables. The ring $R = \mathbb{C}[UV]$, generated over $\mathbb{C}$ by the entries of the product matrix can be thought of in several ways. If $X$ is an $r \times s$ matrix of new variables, then the mapping of $\mathbb{C}[X_{ij}] \to R$ by $X_{ij} \to (UV)_{ij}$ has kernel $I = I_{t+1}(X)$, generated by the $t+1$ size minors of $I$. Thus $R$ is isomorphic to $\mathbb{C}[X_{ij}]/I$. The weaker statement that $I$ is the nilradical of $I_{t+1}(X)$ is equivalent to the statement that the rank of the product of an $r \times t$ matrix with a $t \times s$ matrix is at most $t$. On the other hand, $R$ is also the ring of invariants of $G = GL_t(\mathbb{C})$ acting on $S = \mathbb{C}[U, V]$ via the action $U \to UA^{-1}, V \to AV$. The theorem of Hochster and Roberts implies $R$ is Cohen-Macaulay.

When Emmy Noether in 1921 gave a wonderfully simple proof of Lasker's theorem, as well as establishing an axiomatic treatment of Dedekind domains and rings in which every ideal is finitely generated (now called Noetherian rings), commutative algebra did not exist as a separate field. Indeed some say that it was Noether's paper which ushered in all of modern algebra. It was not until the work of W. Krull (1935) that the field began its process of self-realization. With his principal ideal theorem, and his theory of local rings and their completions, commutative algebra became a distinct entity.

A commutative ring is *local* if it has a unique maximal ideal and is Noetherian. If $R$ is any commutative algebra, and $p$ is a prime ideal, then by forming "quotients" with denominators not in $p$, one may construct a local ring $R_p$. This process is called *localization*. If $R$ is local with maximal ideal $m$, then the powers of $m$ may be used as a neighborhood basis for a topology (which is Hausdorff), and we may complete $R$ with respect to this topology. The result (denoted $\hat{R}$) is again a local commutative ring. For example, if $k$ is a field and $R$ is $k[X_1, \dots, X_n]$ localized at $(X_1, \dots, X_n)$, then $R$ consists of rational functions $f(X_1, \dots, X_n)/g(X_1, \dots, X_n)$ with $g(0, \dots, 0) \neq 0$, while $\hat{R}$ is exactly the ring $k[[X_1, \dots, X_n]]$ of formal power series.

What have we gained by this process? A remarkable amount, as it turns out. In 1946, I. S. Cohen showed that if $R$ is a local ring containing a field, then $\hat{R}$ is isomorphic to a quotient, $k[[X_1, \dots, X_n]]/I$, of a formal power series ring. With this result a standard technique was born: localize, complete, and use the structure of the power series ring to analyze whatever problem one has. We will give an example of this process in a moment.

It became obvious that one type of local ring was the most basic—that of regular local rings, which generalize the classical notion of non-singularity. Examples of regular local rings include power series rings over fields, discrete valuation rings, and the localization of polynomial rings at any prime. In fact, if $R$ is a local ring containing a field, then $R$ is regular if and only if $\hat{R}$ is isomorphic to a power series ring over a field.

A problem which became increasingly famous as its solution evaded researchers again involved the unique factorization property. The conjecture stated that any regular local ring $R$ has the unique factorization property. If $R$ contains a field, then a proof can be given by showing that $\hat{R}$ (a formal power series ring) has unique factorization using the Weierstrass Preparation Theorem, then proving by descent that $R$ also has this property. The proof of the general conjecture, due to M. Auslander and D. Buchsbaum in 1959, ushered in the present era of commutative algebra characterized by the extensive use of homological algebra.

Loosely speaking, homological algebra is linear algebra over an arbitrary ring, or alternatively the study of matrices whose entries lie in an arbitrary ring. One famous problem should convince the reader this is highly nontrivial. If $f_1, \dots, f_n$ are polynomials in $R = k[X_1, \dots, X_m]$ which form the first row of an invertible $n \times n$ matrix with coefficients in $R$, then Cramer's rule shows there are polynomials $g_1, \dots, g_n$ such that $\sum_{i=1}^{n} g_i f_i = 1$. J. P. Serre asked if the converse was true, which was finally proved independently by Quillen and Suslin in 1976. The simplest proof now available uses the technique of localization. (A better known version of Serre's question asked if every projective module over a polynomial ring is free.)

Homological algebra investigates the relations (syzygies) of an arbitrary matrix. The module of relations is finitely generated (Hilbert) and we may put the generators into a matrix and ask for their relations (second syzygies). Hilbert's syzygy theorem states that if $R$ is a polynomial ring, this process stops in finitely many steps! The starting point of the modern viewpoint was the realization that this property *characterizes* regular local rings.

A great deal of modern commutative algebra has been directly inspired by considerations in algebraic geometry, as in the work of Serre and Grothendieck. However, by no means is the entire field a subject of this benevolent tyrant. Newly found connections with topology and combinatorics as well as a deeper self-analysis characterize an expanding field. While the variety of examples in Harry Hutchins' book mirrors the diversity now found in commutative algebra, there are many other important examples which could have been included. Indeed, one of the main pedagogical problems confronting us today is a lack of agreement over what is truly basic in the field.

*Fair Representation*: *Meeting the Ideal of One Man*, *One Vote*. By Michel L. Balinski and H. Peyton Young. Yale University Press, New Haven, CT. 1982. xi + 191 pp., $27.50.

WILLIAM F. LUCAS
*Center for Applied Mathematics, Cornell University, Ithaca, NY 14853*

The apportionment problem deals with the routine operation of rounding fractions, when they are constrained to sum to a given constant. For example, when rounding off numbers in a statistical table so that they sum to precisely 100.0%; when a dean rounds salary increments to be multiples of $100; or when a department chairperson uses the preregistration figures to determine how many sections there should be for each subject. This common procedure appears, at first glance, to be a trivial one in practice, and to warrant little serious concern; whereas the scheme used to apportion fractions may, in some instances, have major consequences. For example, the President of the United States was determined in 1876 by the apportionment used for the U. S. Congress in the 1870s, which was not even in agreement with the method they were supposed to be using at the time.

The authors have chosen to discuss the apportionment problem in the context of fair political representation, and they relate the fascinating history it has had in the U. S. House of Representatives. Great American statesmen like Hamilton, Jefferson, Madison, Washington, J. Q. Adams, and Webster were involved with this mathematical problem and the frequently recurring debates over apportionment of the U. S. House. In the twentieth century, several famous mathematical scientists also took part, as indicated by the long time feud between E. V. Huntington of Harvard and W. F. Willcox of Cornell. Balinski and Young now argue in support of the method of Webster and Willcox over the method of Hill and Huntington which is the current law. (They no longer advocate the use of their own quota method, which was described in this MONTHLY (vol. 82 (1975) 701–730), since they found that it can take a seat away from the *only* state which has gained in population.) It should be stressed, moreover, that the apportionment problem arises in a great number of other allocation problems. It has been studied for various allocations which arise in a typical school district, for manpower assignments in the U. S. Navy, and for the distribution of public transportation vehicles in Leningrad.

This book is likely to become a definitive contribution to a newly emerging field which can be named "equity analysis." This subject is concerned with the fair division of various types of goods and services, as well as political representation. This theory will naturally divide into two parts: finely divisible commodities (money, time, cakes), and more discrete things (people, votes, calculus classes, vehicle loads). *Fair Representation* is destined to become a classic in the latter direction. Although equity has been studied mostly by economists, for example, Gunnar Myrdal (1898– ) and Arthur Okun (1928–1980), it now appears ripe for accelerated development by mathematical scientists; perhaps similar to the development of subjects such as linear programming, game theory, and parts of operations research, in recent decades.

The extensive use of the axiomatic method in classical applied mathematics has been seriously questioned. See, for example, "Beware of Axiomatics in Applied Mathematics" by L. C. Woods, *Bulletin of the Institute of Mathematics and its Applications* (vol. 9 (1973) 41–44). There is surely some value and efficiency in axiomatizing subjects, say in the physical sciences, *after* they have matured somewhat. One unnamed colleague referred to this as "like cleaning up the workshop after the inventor has gone home." There is, nonetheless, some risk to teaching a scientific subject in a way as to downplay the importance of intuition, experimentation, and the modelling process. In contrast to these latter views, axioms do play a very crucial role in the *early* stages of modelling in many of the contemporary directions of applied mathematics, such as in the social, managerial and decision sciences. The book under review is a truly splendid illustration of this latter use of axioms, where intended goals or desirable properties are listed as axioms to be satisfied whenever possible. Over the past decade, Balinski and Young have provided multiple characterizations of the classical methods of apportionment, including axiomatic definitions. It is interesting to note

that great mathematicians of the likes of Huntington and von Neumann reviewed the work on apportionment, but did not bring the axiomatic approach to bear on this problem.

This book also provides a beautiful example of an "impossibility theorem" which can be lucidly explained, and possibly fully proved, to college freshmen. Three very intuitive and highly desirable properties, which have been well known in apportionment considerations for more than 100 years, are in general inconsistent. No apportionment method can always satisfy the quota condition, population monotonicity, and monotonicity in house size (that is, avoid the "Alabama paradox").

In short, this is a delightful book to read, while being a major synthesis of an important area. It is well suited for illustrating some significant points which should be well known to mathematics students as well as to various social scientists.

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

The proof given by J. A. Gallian (this MONTHLY, 91, p. 134) that $A_5$ is simple is indeed simple. However, it can be made even simpler by proving the case $|N| = 2$ as follows.

If $|N| = 2$, then $N$ must be a cyclic group generated by an element of order 2 which must be of the form $(ab)(cd)$. Since $N$ is normal in $A_5$, it must contain the element

$$\begin{aligned} &(abc)^{-1}(ab)(cd)(abc)\\ &=(acb)(ab)(cd)(abc)\\ &=(ad)(bc)\notin N, \end{aligned}$$

which is a contradiction.

K. L. Teo
M. D. Hendy
Dept. of Mathematics and Statistics
Massey University
Palmerston North, New Zealand

Editor:

A number of readers of the MONTHLY have observed that my proof of the simplicity of $A_5$ which appeared in the February 1984 issue could be shortened by using a permutation argument to handle the last case considered. I thank these readers for their comments.

Unfortunately, many readers missed the point of my note. My purpose was to provide a proof which utilizes the same methods to prove that $A_5$ is simple (Sylow theorems and counting techniques) as are typically used to prove the non-existence of simple groups of various orders. Beyond being a nice application of these methods, such a proof has the advantage that it proves

$A_5$ is simple, independent of the five forms in which it commonly arises: as a group of permutations, as the rotation group of a dodecahedron or an icosahedron, as a group of $2 \times 2$ matrices over a field of order 4 or 5 modulo the scalar matrices.

Joseph A. Gallian
Department of Mathematical Sciences
University of Minnesota
Duluth, MN 55812

---

## 145. MISCELLANEA

### Ode to a $\phi$?

A thing as lovely as a $\phi$
I think that I shall never $\left\{\begin{matrix}\text{see}\\ \text{spy}\end{matrix}\right\}$. [Choose one.]

If you want to make $\phi$ rhyme with "see",
But you're really determined to be
  Consistent, admit
  Though it doesn't quite fit,
That a disk measures $r^2$ times "pea".

Each of $\xi$, $\pi$, $\phi$, $\chi$, $\psi$ will claim
An identical rhyme for its name.
  Before you get hot,
  Check Liddell and Scott[1];
Give them, not the author, the blame.

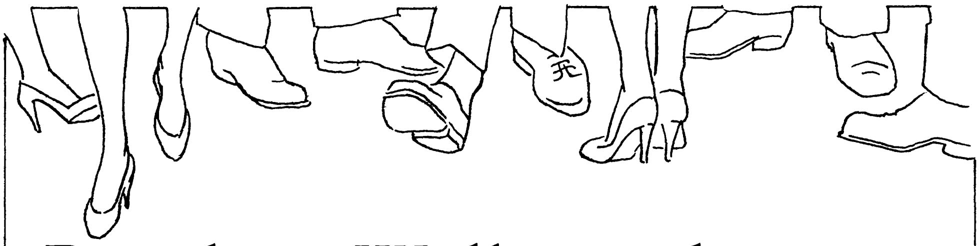
Barry W. Brunson
Department of Mathematics
Western Kentucky University
Bowling Green, KY 42101

[1]A Greek-English Lexicon, compiled by H. G. Liddell & R. Scott, Oxford at the Clarendon Press, 1968 (1st edition 1843; widely recognized as authoritative).

---

### ANSWER TO PHOTO ON PAGE 269

A. N. Kolmogorov



Publications by the

**Mathematical Association of America**

1529 Eighteenth St., NW
Washington, DC 20036
202 387-5200

## New Mathematical Library #30 . . .

**The Role of Mathematics in Science**
by Leon Bowden and M.M. Schiffer
220 pp. Paperbound
List: $16.00 MAA Member: $12.50

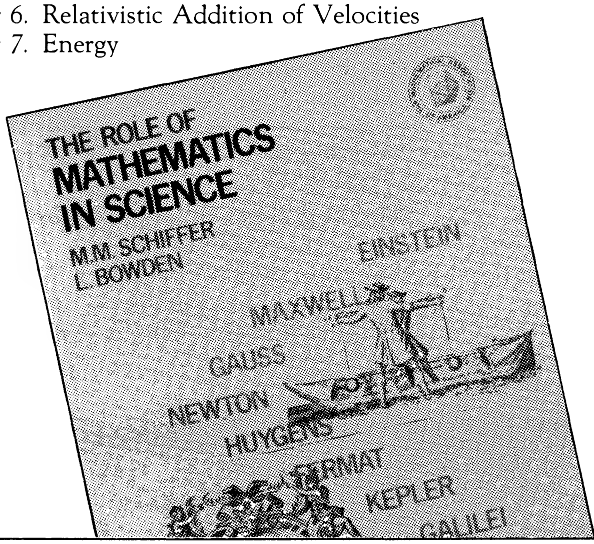
This book is based on a series of lectures given over a period of years to high school mathematics teachers. The lectures focused on topics in mechanics, questions of population growth, probability and other uses of exponential functions, optics and application of matrices to relativity theory.

The aim of the book is to illustrate the power and elegance of mathematical reasoning in science with some examples ranging from the work of Archimedes to that of Einstein. The book starts with problems of the lever, the mirror and the growth of populations and ends up with problems of space travel and atomic energy.

This excellent book will appeal to the advanced high school student, to the undergraduate and to their teachers. A sampling of the Table of Contents should inspire you to order your copy now.

**Table of Contents**





# From The Mathematical Association of America

**VOLUME 6—DOLCIANI MATHEMATICAL EXPOSITIONS**

## MAXIMA AND MINIMA WITHOUT CALCULUS,

by Ivan Niven
xv + 323 pp. Hardbound
List: $28.50  MAA Member: $21.30

Ivan Niven, distinguished author of several books on number theory and probability, has compiled the basic elementary techniques for solving maxima and minima problems. Since many books and courses already cover calculus and linear programming techniques, the author deliberately omits these areas from his discussions and concentrates instead on methods in algebra and geometry not so widely known. These methods are organized according to the mathematical ideas used, and many chapters can be read independently without reference to what precedes or follows. Some of the problems presented are left for the reader to solve with sketches of solutions given in the later pages.

The book is written for an audience at or near the maturity level of second- or third-year college students with a good working knowledge of precalculus mathematics. Although calculus is not a prerequisite, a prior knowledge of that subject will enhance the reader's comprehension. An excellent sourcebook in the area of maxima and minima, the book will serve as a textbook or as enrichment material for the talented undergraduate.

The main topics covered are:

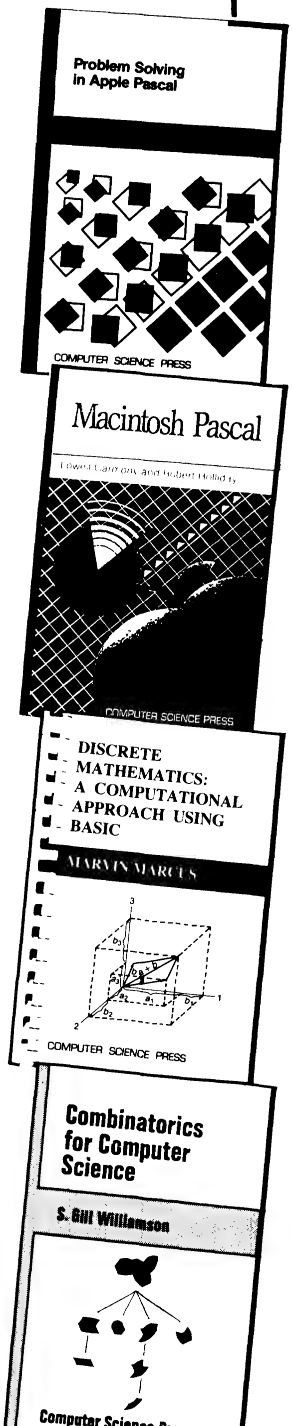
- **Simple Algebraic Results**
- **Elementary Geometric Questions**
- **Isoperimetric Results**
- **Basic Trigonometric Inequalities**
- **Polygons Inscribed and Circumscribed**
- **Ellipses**
- **The Bees and Their Hexagons**
- **Further Geometric Results**
- **Applied and Miscellaneous Problems**
- **Euclidean Three-Space**
- **Isoperimetric Results not Assuming Existence**
- **Postscript on Calculus**

THE MATHEMATICAL ASSOCIATION OF AMERICA

*Order from:*
**The Mathematical Association of America**
1529 Eighteenth St. NW
Washington, DC 20036

# COMPUTER SCIENCE PRESS

## for quality books on math and computers




**Computer Science Press** publishes an outstanding **Computers and Math Series** which focuses on integrating the use of computers and the study of mathematics. **Dr. Marvin Marcus,** editor of the series, is a highly acclaimed mathematician and computer expert.

**MACINTOSH PASCAL,** ***Lowell Carmony and Robert Holliday***
If you have the Macintosh® computer and want to teach your students Pascal, this is the book you need. A convenience diskette* and a solution diskette** are also available. *1985. Book, 0-88175-081-6, $19.95 paper; convenience diskette, 0-88175-088-3, $20.00; solution diskette, 0-88175-089-1, $15.00.*

**DISCRETE MATHEMATICS: A Computational Approach Using BASIC,** ***Marvin Marcus***
This book helps you use a computer creatively and effectively by providing the basic mathematical concepts necessary. A convenience diskette* and a solution manual** are also available. *1984. 0-914894-38-2, $19.95 paper; convenience diskette for the Apple Computer, 0-88175-001-8, $17.00; solution manual, 0-88175-005-0.*

**PROBLEM SOLVING IN APPLE PASCAL,** ***L. A. Carmony, R. J. McGlinn, A. M. Millman, and J. P. Becker***
If you have an Apple® computer, this is an excellent introduction to programming in Pascal on it. A convenience diskette,* teacher's guide,** and solution diskette** are also available. *1984. Student's book, 0-88175-006-9, $19.95; student's diskette, 0-88175-022-0, $15.00; teacher's guide, 0-88175-023-2, $15.00; teacher's diskette, 0-88175-023-9, $17.00.*

**COMBINATORICS FOR COMPUTER SCIENCE,** ***S. Gill Williamson***
Enumeration and graph theory, the two major subdivisions of combinatoric mathematics for computer science, are covered. Emphasis is on the conceptual needs of computer science. *1985 0-88175-020-4, $39.95.*

* *Contains all the programs in the book and saves time keyboarding. You may make copies for students in courses using the book.*
** *Available to instructors using the book as an adopted course text.*

**For a complete description of these and other quality CSP titles, send for a free catalog today.**

*All prices subject to change without notice. Postage and handling is $2.00 for the first book, $1.00 for each additional book. All individual orders must be prepaid. Add appropriate state and local taxes.*

***For prepaid orders:***
**COMPUTER SCIENCE PRESS,** Dept MM385, P.O Box 1659, Hagerstown, MD 21740 USA

***For further information or a free catalog:***
**COMPUTER SCIENCE PRESS,** Dept MM385, 11 Taft Court, Rockville, MD 20850 USA, (301) 251-9050

Macintosh is a trademark of Apple Computer, Inc.
Apple is a trademark of Apple Computer, Inc

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 5 May 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.




The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# ADVICE ON HANGING PICTURES

F. J. BLOORE
*Department of Applied Mathematics and Theoretical Physics, The University, P.O. Box 147, Liverpool L69 3BX, England*

H. R. MORTON
*Department of Pure Mathematics, The University, P.O. Box 147, Liverpool L69 3BX, England*

**1. Introduction.** The irritating way in which pictures frequently hang askew can be explained by simple mechanics related to the geometry of the picture and its cord.

Suppose that the picture is suspended by a cord attached to its two upper corners, passing over a frictionless hook in the wall. Assume also that no friction acts between the picture and the wall. The picture will hang in equilibrium with its centre vertically below the hook; the parts of the cord from the hook to the two corners will make the same angle with the vertical.

We shall see shortly that if the cord is shorter than some critical length, the symmetric equilibrium position is unstable. The picture will, if slightly disturbed, fall askew on one side or the other to a stable but slanting equilibrium position. This situation is illustrated in Fig. 1 where the cord lengths are the same in all three pictures but the tilted pictures have a lower centre of gravity.

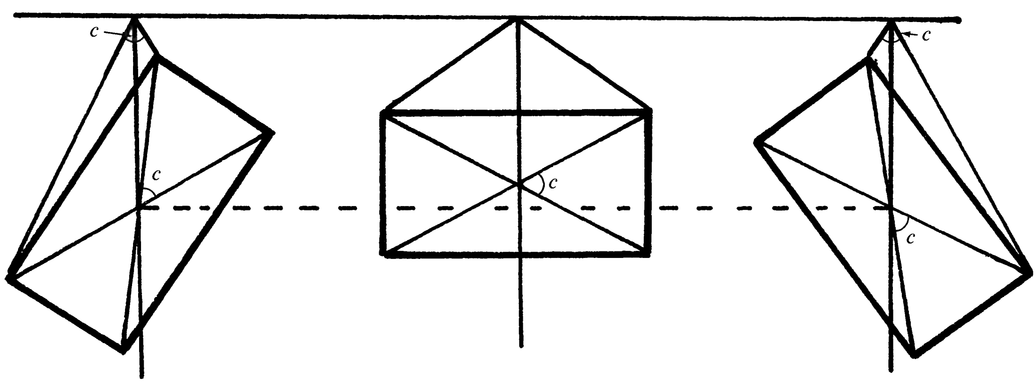



FIG. 1

It turns out (Section 2) that in the tilted equilibrium positions, the angle between the cords—the "support angle"—is equal to the angle $c$ between the diagonals of the picture. The hook then lies on the circle $\mathscr{S}$ through the centre of gravity and the two corners $A$, $B$ where the cord is attached to the picture (see Fig. 2). If in the symmetrical level position of the picture the cord is so short that the support angle is larger than $c$, then this position is unstable. The cord must be lengthened to reduce the support angle to $c$ in order to attain stability of the level

*Fred Bloore*: I took a Ph.D. in nuclear theory in 1961. My interests later drifted towards geometrical aspects of things, particularly of quantum mechanics but also of more concrete objects such as beach pebbles and sand dunes. I enjoy hill walking and am a fairly active member of Amnesty International.

*Hugh Morton*: As a research student at the University of Warwick in 1964–66, I saw the very beginnings of the Mathematics department there, under the inspiration of Chris Zeeman; since then I have been a lecturer in Pure Mathematics at Liverpool. My general enjoyment of things geometrical is normally directed at low-dimensional topology, particularly knot theory and other areas in which I can try to visualise the objects under study.

By way of relaxation I enjoy making geometrical models from time to time. Fred and I hope to develop a robust demonstration model from the ideas in this article, to take its place among other mathematically based exhibits in our departmental display.

position. The hook now lies on the perpendicular bisector $\mathscr{L}$ of $AB$.

The problem is acute for a long thin picture such as a school photograph or a panorama. These have a small angle $c$ so need a long cord. As an extreme example, if the Bayeux tapestry ($\frac{1}{2}$m $\times$ 70m) were nailed down on a rigid rectangle of hardboard, one would need 1 km of cord to hang it so the horizontal equilibrium position were stable!

By altering the points of attachment of the cord we can change the effective geometry of the picture. The important features are the position of the centre of mass of the picture, and the two points of attachment. Provided that both points of attachment lie at the same distance from the centre of mass, the picture will behave as if it were a rectangular picture with the same centre of mass, whose corners lie at the points of attachment. The effective characteristic angle $c$ can thus be increased by moving the points of attachment closer to each other along the top edge of the picture.

In the next two sections we shall show why the equilibria follow the patterns described, and analyse their stability. As the cord length is varied, the equilibrium position of the picture moves relative to the hook. It is fruitful to consider the *locus of the hook relative to the equilibrium position of the picture* as the cord length is varied. This locus is the union of the line $\mathscr{L}$ and the circle $\mathscr{S}$, Fig. 2. As the cord is shortened, the hook moves down $\mathscr{L}$ to $E$, then along one or other arc of $\mathscr{S}$. The continuation of $\mathscr{L}$ below $E$ is unstable.

The main phenomenon is the splitting of one stable equilibrium position into two as the cord is shortened (or as the cord is kept the same length and the centre of gravity raised!). The nature of the transition which occurs is familiar from the viewpoint of elementary catastrophe theory. This naturally suggests that we consider also what happens when the centre of gravity is not equidistant from the points of support.

In this case, treated in Sections 4 and 5, the locus of the hook $H$ relative to the picture is no longer a circle and line but a more general cubic curve called a *strophoid*. The points of attachment of the picture also lie on this strophoid. They are conjugate with respect to a natural involution on the strophoid, and we explore the geometry from this point of view.

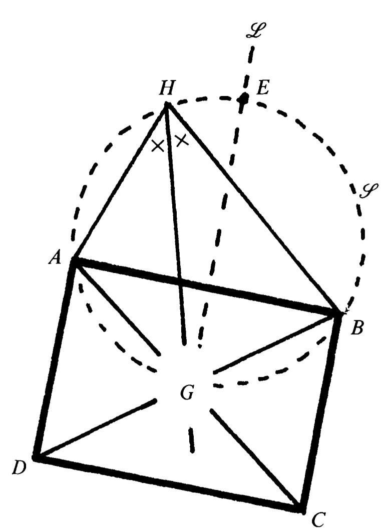



FIG. 2(a). Short cord.

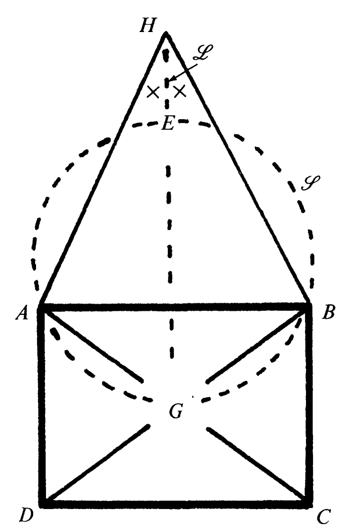



FIG. 2(b). Long cord.

**2. Circle and Ellipse.** Consider a rectangular picture $ABCD$ of length $AB = 2a$, depth $BC = 2b$ with centre of mass $G$ at its centre, suspended by a cord of length $2l$ attached at $A$ and $B$ passing over a hook $H$, Fig. 2.

We assume that no friction acts between the picture and the wall. Then $G$ must lie vertically below $H$ since the three forces on the picture, namely gravity and the tensions in $AH$ and $BH$,

must pass through the same point $H$. The tensions in the two parts of the cord are equal, because the hook is frictionless. Their horizontal components must cancel, and so the pieces of cord $AH$ and $BH$ make equal and opposite angles with the vertical $HG$. It follows that *$H$ must lie on either*

(1) the perpendicular bisector $\mathscr{L}$ of $AB$, *or*
(2) the circle $\mathscr{S}$ which passes through $A$, $G$ and $B$.

In the latter case $H$ must be above the arc $AGB$, and the equality oı the angles is implied by the equality of the chord lengths $AG$ and $GB$. We now prove that $\mathscr{L}$ and $\mathscr{S}$ are the only possible locations for $H$. Let $B'$ be the reflection of $B$ in the line $HG$, Fig. 3.

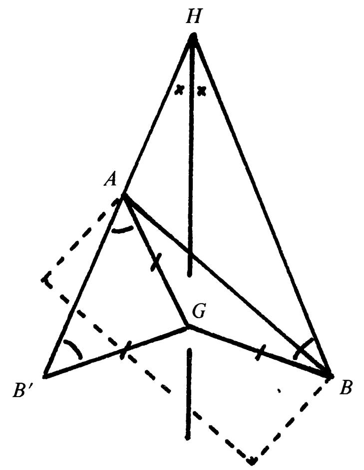



FIG. 3

Since the angles $AHG$, $BHG$ are equal, $B'$ must lie on $HA$.

If $B' = A$, then $HG$ is perpendicular to $AB$ and so $H$ lies on the perpendicular bisector of $AB$. If $B' \neq A$, then

$$\begin{aligned} HBG &= HB'G \text{ by reflection} \\ &= B'AG \text{ since the triangle } B'AG \text{ is isosceles} \\ &= \pi - HAG. \end{aligned}$$

Thus $HAGB$ is a cyclic quadrilateral; $H$ lies on the circle through $A, G, B$. Note that when $H$ lies on the arc $AGB$ of $\mathscr{S}$, the angles $AHG$, $GHB$ are not equal but supplementary.

Since $\mathscr{S}$ and $\mathscr{L}$ are fixed relative to the picture, it is natural to consider the equilibrium positions of the hook $H$ relative to the picture, and how they change as we vary the length of the cord, Fig. 4.

Let $\mathscr{L}$ cross $\mathscr{S}$ at $E$, above $G$. Then $AE = a\sqrt{\left(1 + (a/b)^2\right)}$.

For "long" cords, i.e., for $l \geqslant AE$ the hook must lie on $\mathscr{L}$ above $E$. There is only one position of equilibrium; the picture hangs level in stable equilibrium. For "short" cords $a < l < AE$, there are three positions of equilibrium, two asymmetrical ones $H_1$ and $H_3$ on $\mathscr{S}$ flank a position $H_2$ on $\mathscr{L}$ (Fig. 4). It is well known that for fixed cord length $2l$, the locus of $H$ such that $HA + HB = 2l$ is an ellipse $\mathscr{E}_l$ with foci $A$ and $B$ (and with eccentricity $e = a/l$). For long cords $\mathscr{E}_l$ will not cross $\mathscr{S}$ but will cross $\mathscr{L}$ at $H$. For short cords, $\mathscr{E}_l$ will cross $\mathscr{S}$ at $H_1$ and $H_3$, and will cross $\mathscr{L}$ at $H_2$ (and also at $H_4$), Fig. 4.

The stability of the positions of equilibrium can be decided by examining the potential energy of the picture. The potential energy of the picture relative to the hook is proportional to $-GH$. A position $H$ is an equilibrium position if $GH$ is stationary with respect to movements of $H$ along the ellipse, i.e., if $GH$ is normal to the ellipse. The equilibrium will be stable if $GH$ is a local

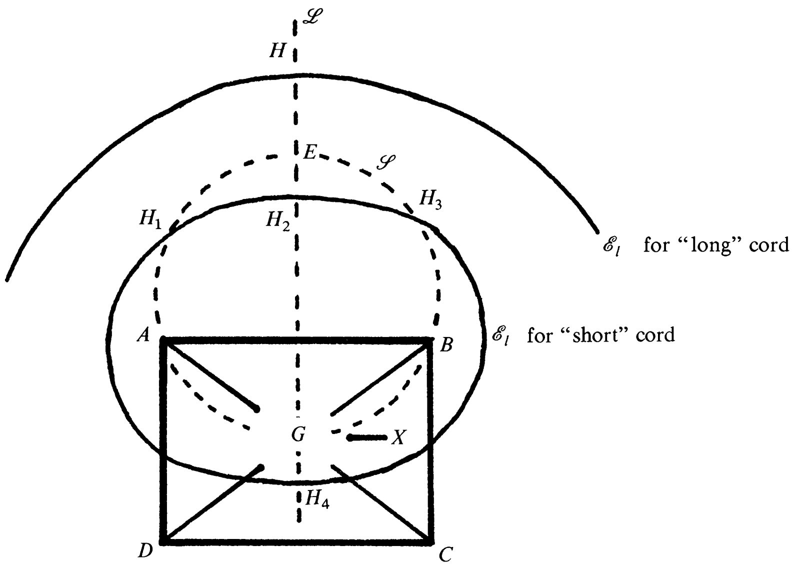



FIG. 4

maximum, unstable if a local minimum.

The circle $\mathscr{C}$ with centre $G$ and radius $GH_1$ touches the ellipse $\mathscr{E}_l$ at $H_1$ and $H_3$. The two curves cannot meet elsewhere since distinct conics can intersect at most four times, and $\mathscr{C}$ and $\mathscr{E}_l$ have double contact at $H_1$ and at $H_3$. So the ellipse $\mathscr{E}_l$ must lie entirely within the circle $\mathscr{C}$, because the point $H_4$ on $\mathscr{E}_l$ is certainly inside $\mathscr{C}$. We conclude that $GH_1$ and $GH_3$ are local (in fact global) maxima and $GH_2$ is a local minimum.

Now draw $GX$ tangent to the circle $\mathscr{S}$. One of the classical circle theorems tells us that angle $XGB$ equals angle $GH_1B$. Thus for rectangular pictures on short cords, the symmetrical position is unstable, and the picture tilts until the angle $AH_1B$ or $AH_3B$ between the cords has decreased down to the angle $BGC$ between the diagonals of the picture. More usefully, *to hang a picture in symmetrical stable equilibrium, the cord must be long enough that the angle between its halves is less than or equal to the angle between the diagonals of the picture*. If the points of attachment $A$ and $B$ of the string to the picture are not at the corners, the criterion becomes: the angle $AHB$ must not exceed $\pi$ minus angle $AGB$. So for a given picture and cord length, one can obtain stability in the

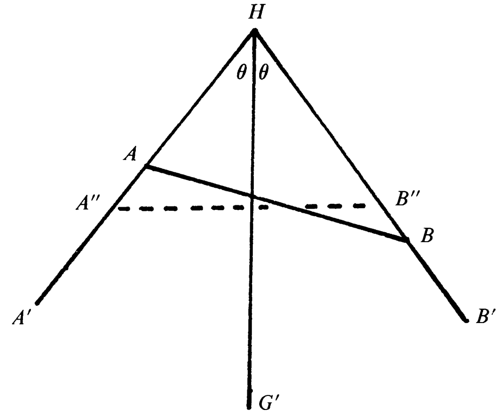



FIG. 5

symmetrical position by moving the points $A$ and $B$ sufficiently close together, but keeping $AG = BG$.

Now return to the rectangular picture hung by the corners, and consider its equilibrium positions *relative to the hook H on the wall*, as the cord increases in length from $2a$ (Fig. 5). Let the angle $BGC$ between the diagonals be $2\theta$; draw a line $HG'$ on the wall vertically downwards and lines $HA'$, $HB'$ at an angle $\theta$ to $HG'$. Then the points $A$ and $B$ of the picture will lie on $HA'$ and $HB'$ and the point $G$ on $HG'$. As the length of the cord increases, the picture will move downwards and become eventually horizontal, when the cord reaches its critical length. The circle $AGBH$ moves with the picture and *rolls* on a circle twice as large on the wall, centre $H$. This is because $A$, $B$ and $G$, respectively, traverse the fixed radii $HA'$, $HB'$, $HG'$ of the big circle on the wall, Fig. 6. These radii are perpendicular to the lines $PA$, $PB$, $PG$, where $P$ is the point of contact of the two circles. Hence $P$ is the instantaneous centre of rotation of $\mathscr{S}$. If the cord is lengthened further, the picture remains horizontal and moves downwards; $A$ and $B$ leave the lines $HA'$, $HB'$.

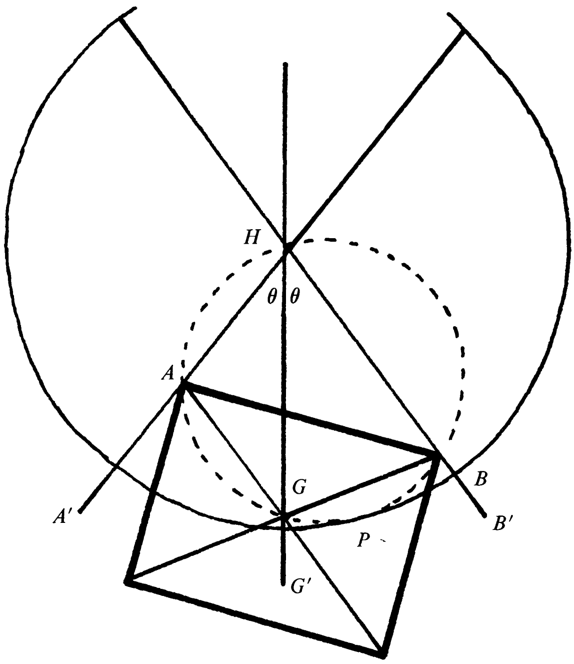



FIG. 6

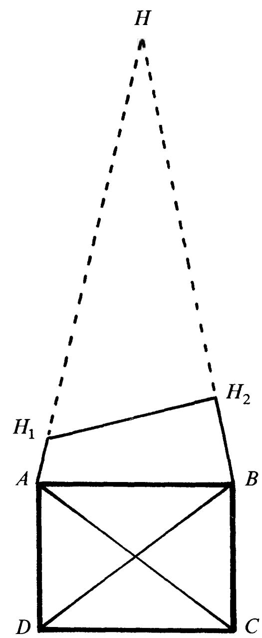



FIG. 7

It must be admitted that one could "cheat" and use two hooks. In Fig. 7, the picture behaves as if suspended from $H$ whereas in reality a much shorter cord $AH_1H_2B$ is being used. The points $H_1$, $H_2$ may be chosen anywhere on the lines $HA$ and $HB$. We invite readers to develop this case further. (We haven't.)

**3. Evolute.** We return to the viewpoint relative to the picture, and present another approach to the stability question using the evolute of the ellipse $\mathscr{E}_l$ of Fig. 4. This is the locus of the centre of curvature of $\mathscr{E}_l$, which is an astroid-like curve $\mathscr{A}_l$, Fig. 8. If $G$ lies inside $\mathscr{A}_l$, i.e., if the lower cusp $J$ of $\mathscr{A}_l$ lies below $G$, then the circle of centre $G$ and radius $GH_2$ will be a tighter circle than the circle of curvature at $H_2$ (which has centre $J$ and radius $JH_2$). Hence $GH_2$ will be a minimum of $GH$ as $H$ moves on the ellipse and so $H_2$ is an unstable position of equilibrium.

If the cord is lengthened, the ellipse $\mathscr{E}_l$ becomes larger and more nearly circular; the curve $\mathscr{A}_l$ becomes smaller, shrinking down to the centre point $O$. Its lower cusp $J$ will move upwards

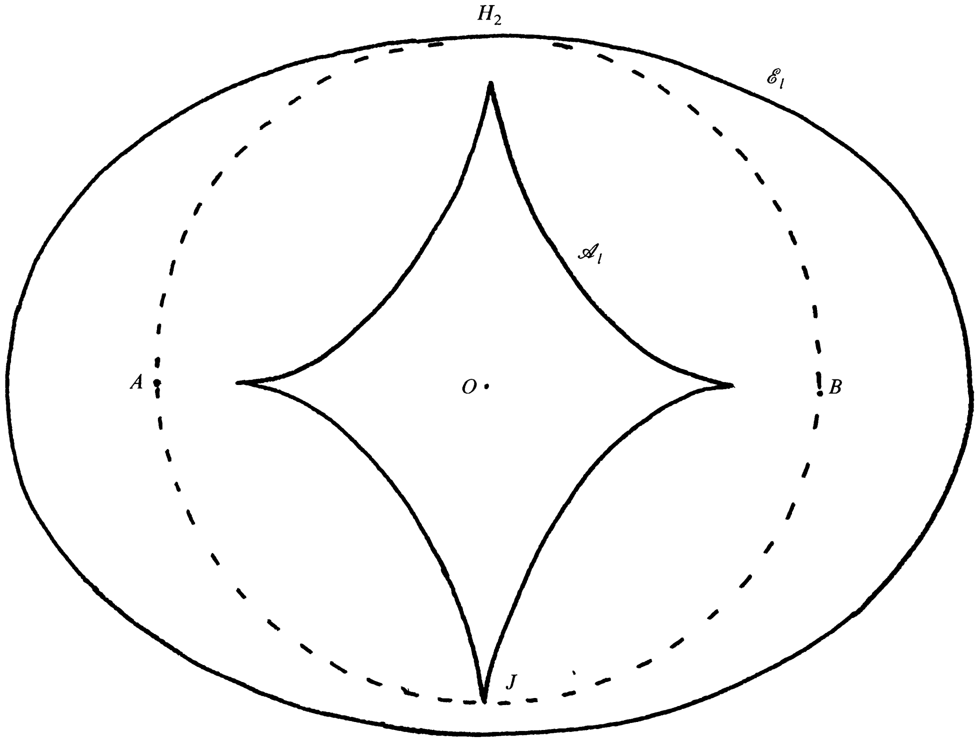



FIG. 8

through $G$ towards $O$. At the critical length of cord when $J$ is on top of $G$, the three equilibrium positions $H_1, H_2, H_3$ (which we observed to lie on the circle through $A$, $B$ and $G$) must coincide, corresponding to the (well-known) theorem that for any ellipse the foci $A$ and $B$, the point $H_2$ and its centre of curvature $J$ lie on a circle whose radius is half the radius of curvature at $H_2$.

If the string is long enough that $J$ is above $G$, then the circle of curvature at $H_2$ will be tighter than the circle of centre $G$ and radius $GH_2$, so that $H_2$ is a stable position of equilibrium.

**4. Catastrophe theory.** Readers acquainted with catastrophe theory will recognise that the cusp $J$ which governs the stability character of the hanging picture arises from a cusp catastrophe. We now develop this aspect. Several clear introductions to catastrophe theory exist already (see the introductory chapter of reference **2**) so we shall just sketch the main idea.

Consider a potential function $V(\alpha_1, \dots, \alpha_m, x_1, \dots, x_n)$ of a system having coordinates $x_1, \dots, x_n$ and adjustable parameters $\alpha_1, \dots, \alpha_m$. Catastrophe theory describes how the stationary points of $V$ (solutions of $\partial V/\partial x = 0$) can coalesce and disappear as the parameters $\alpha$ are varied.

The simplest example is $V(\alpha, x) = x^3 + 3\alpha x$. For negative $\alpha$, there are two stationary points $x = \pm\sqrt{-\alpha}$, one stable and one unstable equilibrium, which coalesce and disappear as $\alpha$ increases through the value zero, Fig. 9.

The graph of the stationary points as functions of $\alpha$, $x = \pm\sqrt{-\alpha}$ is the parabola Fig. 10. This sudden loss of stationary points caused by a continuous change of parameter is called a catastrophe.

The next most simple example is

$$V(\alpha, \beta, x) = x^4 - 6\beta x^2 + 8\alpha x. \tag{1}$$

The stationary points are given by $\partial V/\partial x = 0$, i.e.,

$$4x^3 - 12\beta x + 8\alpha = 0. \tag{2}$$

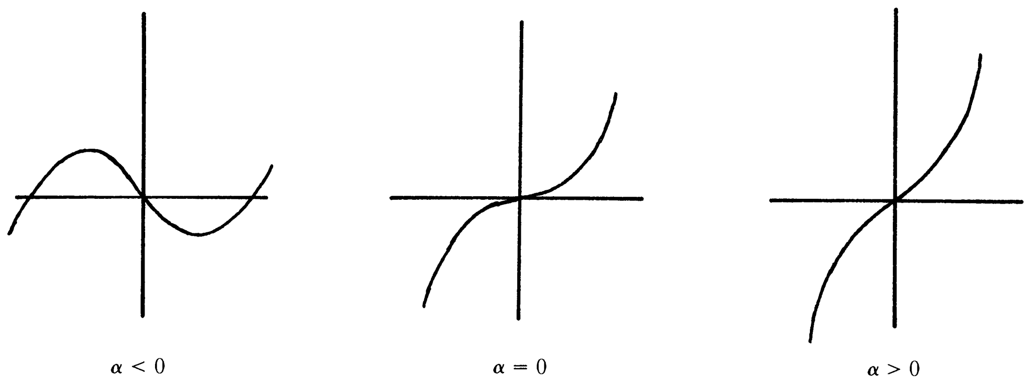



FIG. 9

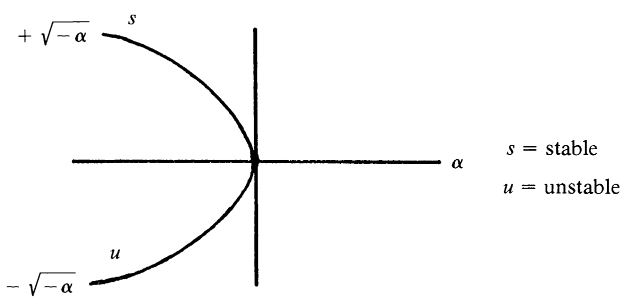



FIG. 10

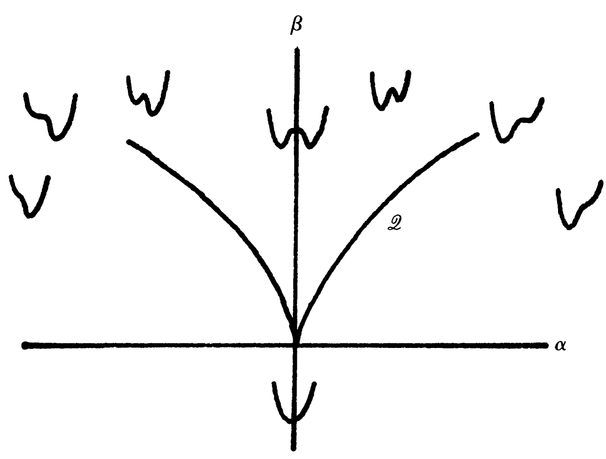



FIG. 11

Two stationary points coincide if also $\partial^2 V/\partial x^2 = 0$, i.e.,

$$(3) \qquad 12x^2 - 12\beta = 0.$$

Eliminating $x$ from (2) and (3) gives $\alpha^2 = \beta^3$, a cubic parabola $\mathscr{Q}$ in $(\alpha, \beta)$-space, Fig. 11.

The small wiggly graphs dotted about on Fig. 11 are little sketches of the graph of $V(\alpha, \beta, x)$ as a function of $x$, for the values $(\alpha, \beta)$ given by the location at which the small graph appears. The function $V$ has three stationary points if $(\alpha, \beta)$ is above the curve $\mathscr{Q}$ ("inside the cusp"). As

$(\alpha, \beta)$ crosses $\mathcal{Q}$, so two of the stationary points of $V$ coalesce and disappear. The local situation is cubic there, as in Figs. 9 and 10.

We now show that the potential energy function (squared, actually) of our hanging picture is, neglecting terms $x^5$ and higher, of the form (1) with $(\alpha, \beta)$ proportional to the coordinates of the centre of gravity $G$, referred to the cusp $J$ of Fig. 8. This is the first time in this article that we allow $G$ to deviate from the perpendicular bisector of $AB$, but we do so from now on. When $G$ is just inside the cusp $J$ of Fig. 8 the picture has 3 equilibria; the hook positions are the feet of the 3 possible perpendiculars drawn upwards from $G$ to the ellipse. (A fourth may be drawn downwards.) As $G$ crosses the evolute $A_l$, two of these feet coalesce and disappear. So the cusp $J$ is a catastrophe cusp like Fig. 11.

Choose axes at $O$, fixed in the picture, with the $x$-axis along $OB$, Fig. 8. Let the hook $H$ have coordinates $(x, y)$. It lies on the ellipse $\mathcal{E}_l$ with equation

$$\frac{x^2}{l^2} + \frac{y^2}{l^2 - a^2} = 1. \tag{4}$$

It follows from coordinate geometry that the cusp $J$ is at $(0, -a^2/(l^2 - a^2)^{1/2})$. Suppose that the centre of gravity $G$ has coordinates $(\alpha, -a^2/(l^2 - a^2)^{1/2} + \beta)$, so that $(\alpha, \beta)$ is the position of $G$ relative to $J$. We shall compute the potential energy (squared) $V(\alpha, \beta, x)$ of the picture as a function of the $x$-coordinate $x$ of $H$, where $H$ is restricted to the ellipse (4).

Apart from a factor $mg$ the potential energy is minus the distance $GH$. We choose instead to study the function $-GH^2$ as this has the same stationary points but fewer square roots in its algebraic expression. Thus we choose

$$V(\alpha, \beta, x) = -GH^2 = -(x - \alpha)^2 - \left(y(x) + a^2(l^2 - a^2)^{-1/2} - \beta\right)^2, \tag{5}$$

where by (4)

$$y(x) = (l^2 - a^2)^{1/2}(1 - x^2/l^2)^{1/2}.$$

Let us expand $y(x)$ in powers of $x$, neglecting terms above $x^4$,

$$y(x) \simeq (l^2 - a^2)^{1/2}\left[1 - \frac{x^2}{2l^2} - x^4/8l^4\right]. \tag{6}$$

The approximation is good for small $x$. Substitution of (6) in (5) gives, to order $x^4$,

$$V(\alpha, \beta, x) = Ax^4 - \beta(l^2 - a^2)^{1/2}x^2/l^2 + 2\alpha x + \text{constant},$$

where $A = (a^2 - \beta(l^2 - a^2)^{1/2})/4l^4$. This is of the form (1), if we neglect the $\beta$-dependence of the coefficient $A$. We are justified in this if $\beta$ is small, which is the case when $G$ is near to $J$.

**5. Catastrophoid!** Finally we ask: What is the locus of the equilibrium positions of the hook $H$ relative to the picture as the cord length $l$ is varied, when the points of attachment $A$ and $B$ are *not* symmetrically placed relative to the centre of gravity $G$? In the symmetrical case this locus was the line $\mathcal{L}$ together with the upper arc $AEB$ of the circle $\mathcal{S}$, Fig. 12. In all cases the locus is specified by the equilibrium condition that the angle $AHG$ should equal the angle $BHG$. When $G$ moves off centre, the locus of line plus part of circle deforms to part of a more general cubic curve, called a right strophoid, Fig. 12. The point of intersection $E$ disappears. The special bifurcation of "trident" type at $E$ becomes a more general bifurcation, Fig. 13 (ref. **3**). The upper stable branch remains stable and does not bifurcate, but a pair of stable and unstable equilibria appear, as in Fig. 10.

We sketch the derivation of the equation of the strophoid, which is straightforward coordinate geometry. Given three points $A$, $B$, $G$ in a plane, we seek the locus of the point $H(x, y)$ such that $AHG = BHG$. Let us choose the same axes as in §4, i.e., let $A$ be $(-a, 0)$, $B(a, 0)$ but relabel the coordinates of $G$ to be $(\alpha, \beta)$.

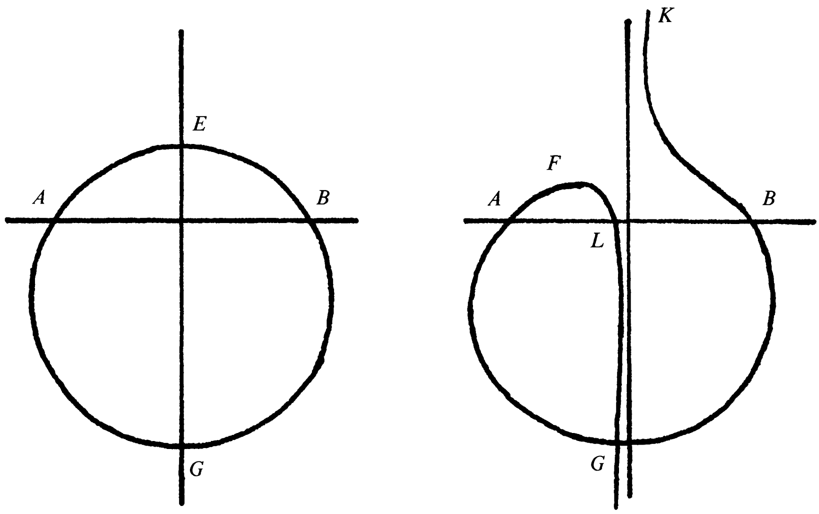



FIG. 12

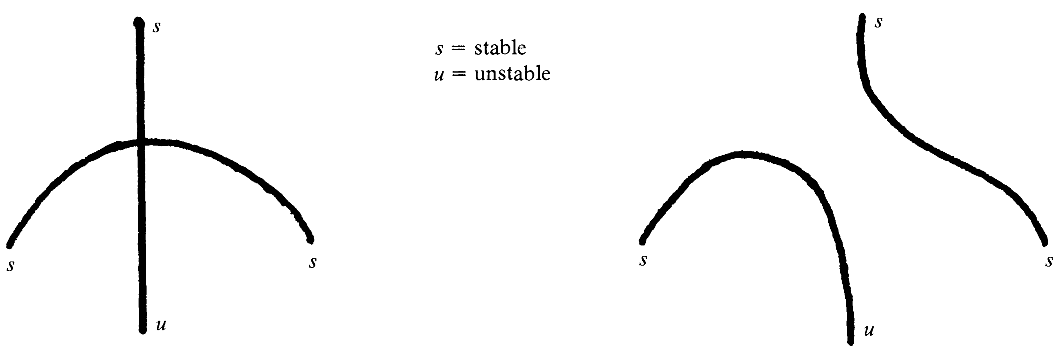



FIG. 13

The condition

$$\cos AHG = \cos BHG$$

becomes, using scalar products,

$$\frac{\mathbf{AH} \cdot \mathbf{GH}}{|\mathbf{AH}|\,|\mathbf{GH}|} = \frac{\mathbf{BH} \cdot \mathbf{GH}}{|\mathbf{BH}|\,|\mathbf{GH}|}.$$

If we cancel $|\mathbf{GH}|$, square both sides, and then cross multiply, we obtain a polynomial equation in $x$ and $y$,

$$\left[(x-a)^2+y^2\right][(x+a)(x-\alpha)+y(y-\beta)]^2$$
$$=\left[(x+a)^2+y^2\right][(x-a)(x-\alpha)+y(y-\beta)]^2.$$

This simplifies to the cubic equation

(7) $$(X^2+Y^2)(\beta X-\alpha Y)+\alpha\beta(X^2+kXY-Y^2)=0,$$

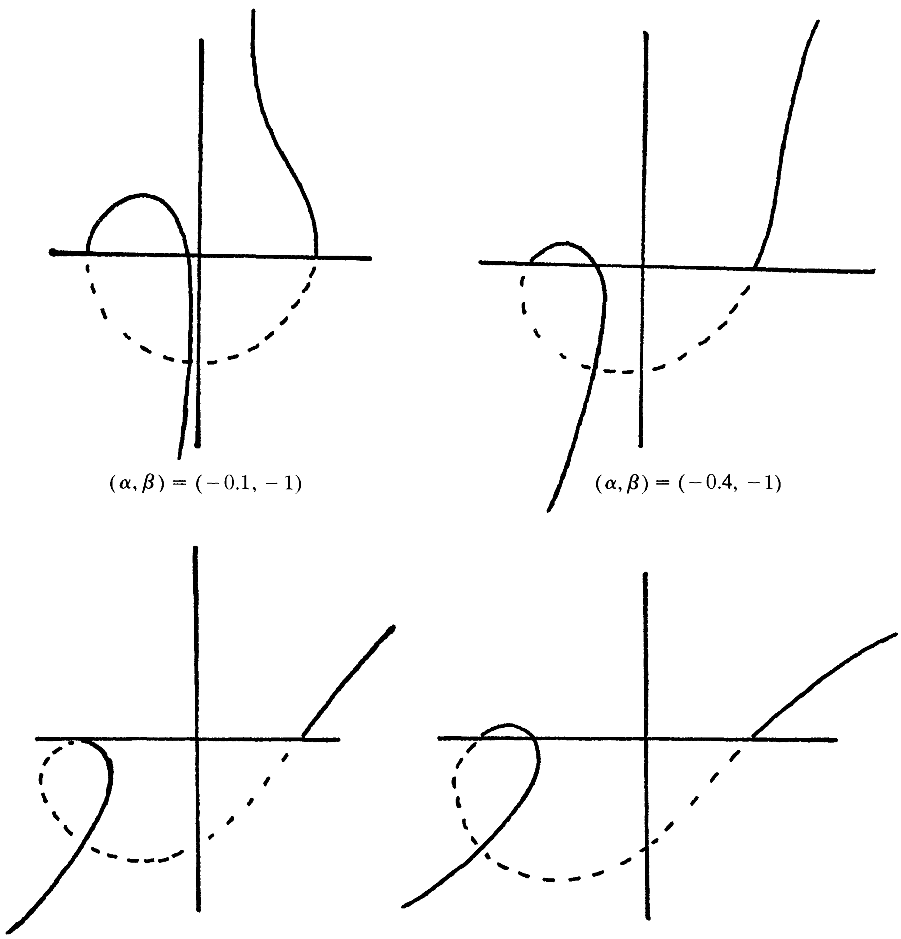
where $X = x - \alpha$, $Y = y - \beta$, and $k = (a^2 + \beta^2 - \alpha^2)/\alpha\beta$, after cancellation of a factor $y$. The possibility $y = 0$ corresponds to the line $AB$ which is obviously part of the locus, but not accessible physically for a picture hook. If $\alpha = \beta = 0$, equation (7) reduces to $xy = 0$. If $\alpha = 0$,

$\beta \neq 0$, equation (7) reduces to

$$x\left\{\beta\left(x^2+(y-\beta)^2\right)+\left(a^2+\beta^2\right)(y-\beta)\right\}=0,$$

the line and circle of §2.

In the general case equation (7) defines a "right strophoid" [**1**], [**4**], which we denote $S(A, B; G)$. It has an asymptote of slope $\beta/\alpha$, a node at $G$ ($X = Y = 0$), at which the curve intersects itself at right angles and it passes through $A$ and $B$. The tangents to the strophoid at $G$ are the bisectors of the angles between $AG$ and $BG$. Some examples are given in Fig. 14, with $a = 1$ and $(\alpha, \beta)$ labelling the curves. Only part of the strophoid corresponds to equality of the angles $AHG$, $BHG$. On the rest of the curve, drawn dotted, the angles are supplementary; this possibility crept in when we squared the cosines.



$(\alpha, \beta) = (-0.1, -1)$ $(\alpha, \beta) = (-0.4, -1)$

$(\alpha, \beta) = (-1, -1)$ $(\alpha, \beta) = (-1.5, -1)$

FIG. 14

The interpretation of Fig. 12 is as follows. For a picture with centre of gravity $G$ and string attached at $A$ and $B$, if the string is long, the equilibrium position $H$ of the hook relative to the picture will lie on the part $BK$ of the strophoid, where the ellipse $\mathscr{E}_l$ cuts the curve. The potential energy function $V = -HG$ or $-HG^2$ has only one stationary point as $H$ moves on $\mathscr{E}_l$. $GH$ is normal to the ellipse $\mathscr{E}_l$.

We make the depressing observation that no matter *how* long the string is, the picture will always hang crooked if the centre of gravity is not equidistant from the points of support! For a "long" cord the centre of gravity will tend to lie vertically below the midpoint of $AB$ since the line through $G$ in the asymptotic direction passes through this midpoint. As the string is shortened the ellipse $\mathscr{E}_l$ shrinks and the hook position $H$ will move down the curve towards $B$, until the ellipse touches the arc $AFL$ of the strophoid, at some point $F$. The potential energy function for this ellipse has a stationary point of inflexion at $F$, and a minimum on the arc $KB$. If the string is shortened further, the ellipse cuts the strophoid at three points, giving stable equilibria on $AF$ and $BK$ moving towards $A$ and $B$, and an unstable one on $FL$ moving towards $L$. The point of inflexion has split apart into a maximum and a minimum. The part of the strophoid below $AB$ is not physically relevant. We see that the strophoid $S(A, B; G)$ *could* be defined to be the locus of the feet of the normals from $G$ to the family of ellipses with foci $A$ and $B$.

In the literature [**1**], [**4**], a strophoid is usually defined in terms of a fixed point $O$ called its centre and a line $\mathscr{L}$ with a point $G$ on it called its node. A point $H$ is on the strophoid if $OH$ intersects $\mathscr{L}$ at $H''$ such that $HH'' = GH''$, Fig. 15.

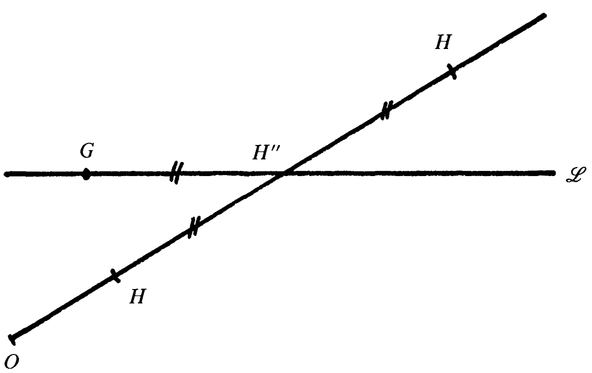



FIG. 15

As $O$ approaches any point $O_0$ on $\mathscr{L}$, so the strophoid degenerates to a circle of radius $O_0G$ and centre $O_0$, together with the line $\mathscr{L}$. The centre of the strophoid of equation (7) is at

$$\left(-\frac{\alpha}{2}\left(1-\frac{a^2}{\alpha^2+\beta^2}\right), -\frac{\beta}{2}\left(1+\frac{a^2}{\alpha^2+\beta^2}\right)\right).$$

The question arises how to reconcile our "equal-angle" construction of a strophoid with the Lockwood construction. How are the points of attachment $A, B$ related to the centre $O$ and the line $\mathscr{L}$? Are there any other pairs of points $C, D$ on the strophoid $S(A, B; G)$ such that the angles $CHG$, $DHG$ are equal or supplementary for each point $H$ on the curve?

The answer to the last question is that *every* point $C$ on the strophoid is paired in this sense to another "conjugate" point $D$, which may be found by drawing the diameter $COC'$ from $C$ through the centre $O$ to the antipodal point $C'$, and then drawing the chord $C'D$ parallel to the nodal line $\mathscr{L}$ to meet the strophoid again in $D$. We see that $G$ is conjugate to itself and $O$ is conjugate to the point on the strophoid at infinity, Fig. 18.

Another construction for $D$ is to draw the two tangents to the strophoid at the node. The lines $GC$ and $GD$ are then mirror reflections of each other in the nodal tangents. But the best construction follows from the direct construction of the strophoid $S(A, B; G)$ now given, Fig. 16, and answers all the questions just raised.

Draw concentric circles centred at $G$ (not enclosing $A$ or $B$). For each circle, draw the two tangents from $A$ and the two tangents from $B$. These intersect in six points, $A, B, H_1, H_2, H_3, H_4$, called the vertices of the complete quadrilateral of the four tangents, Fig. 16. We make two claims.

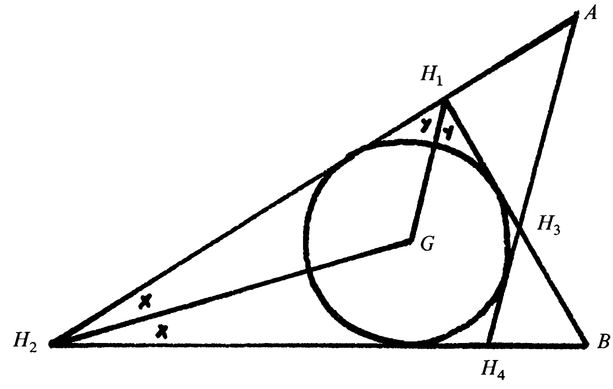



FIG. 16

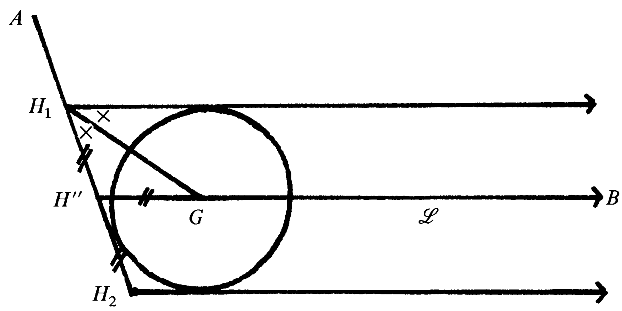



FIG. 17

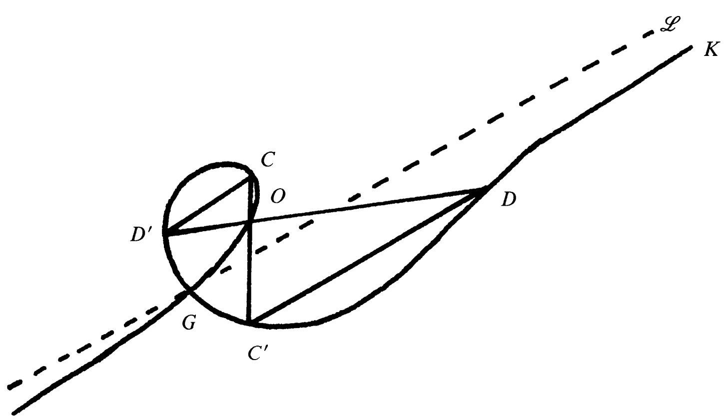



FIG. 18

(i) All these vertices lie on $S(A, B; G)$.

(ii) Opposite vertices are conjugate points for $S(A, B; G)$; that is

$$\begin{aligned} S(A, B; G) &= S(H_1, H_4; G) \\ &= S(H_2, H_3; G). \end{aligned}$$

Claim (i) is obvious from Fig. 16. Note that $AH_2G = BH_2G$ but $AH_1G = \pi - BH_1G$.

As a special case, if we let $B$ go off to infinity along the $x$-axis, Fig. 17, the strophoid $S(A, \infty; G)$ is the same as the strophoid constructed by the Lockwood prescription, with centre $A$,

node $G$ and the $x$-axis as the nodal axis. This is because the tangents from $B$ to all the circles centred at $G$ are parallel. The equality of the angles $x$ in Fig. 17 implies the equalities of the lengths $H''H_1, H''H_2, H''G$.

To prove claim (ii) we consider the two strophoids $S(A, B; G)$ and $S(H_1, H_4; G)$. They have the same six points $A, B, H_1, H_2, H_3, H_4$ in common and they have the same node. Hence they must be the same strophoid, because there is only a three-parameter family of strophoids with a given node $G$. (The three parameters may be chosen to be the direction of the nodal line through $G$, and the two coordinates of the centre $O$.) Hence $H_1$ and $H_4$ are good conjugate points to generate $S(A, B; G)$.

In Fig. 18, if a picture with centre of gravity $G$ is hung from the conjugate points $C$ and $D$ as points of attachment, then the hook must lie on the part $DK$ of the strophoid. If it is hung from the conjugate points, $C'$ and $D'$, then the hook will lie on the same strophoid, but all the strophoid to the right of the line $C'D'$ is now available.

We thank our colleagues Peter Giblin, Chris Gibson, and John Underhill for helpful information and computer pictures.

**References**

1. E. H. Lockwood, A Book of Curves, Cambridge Univ. Press, 1961.
2. P. T. Saunders, An Introduction to Catastrophe Theory, Cambridge Univ. Press, 1980.
3. E. C. Zeeman, Bifurcation, Catastrophe, and Turbulence, pp. 109–153, in New Directions in Applied Mathematics, P. J. Hilton and C. Young, eds., Springer-Verlag, NY, 1982.
4. C. Zwikker, The Advanced Geometry of Plane Curves and Their Applications, Dover, NY, 1963.

## HOW TO SOLVE THE SYSTEM $x' = Ax$

B. VAN ROOTSELAAR
*Heidepark 13, 6705 AB Wageningen, The Netherlands*

**Summary.** A simple formula (7) is derived for the solutions of $x' = Ax$ by reduction to a single equation. This formula is extended to the system $x' = Ax + u$ in conceptually the same way as for a single equation of first order. The method also gives easy constructions for the fundamental matrix, Jordan bases and the Jordan normal form. The method is completely elementary.

**1. Reduction to a single equation.** Consider the system $x' = Ax$, where $A$ is a constant square matrix of order $n$. If the characteristic equation $\det(sI - A) = 0$ of $A$ has the form

$$s^n + a_1 s^{n-1} + \cdots + a_n = 0, \tag{1}$$

then by Cayley-Hamilton's theorem

$$A^n + a_1 A^{n-1} + \cdots + a_n I = 0.$$

Hence for any $x$

$$A^n x + a_1 A^{n-1} x + \cdots + a_n x = 0.$$

If $x$ is a solution of $x' = Ax$, then $x^{(k)} = A^k x$, so any solution of $x' = Ax$ is a solution of the (vector) differential equation

$$x^{(n)} + a_1 x^{(n-1)} + \cdots + a_n x = 0. \tag{2}$$

Let $s(1), \ldots, s(p)$ be the different roots of (1), with multiplicities $m(1), \ldots, m(p)$. To the $m(j)$-fold root $s(j)$ there correspond $m(j)$ linearly independent solutions of a single equation of

(2) of the following form:

$$f_{j,k}(t) = (t^k/k!)\exp(s(j)\cdot t) \qquad (k = 0,\dots,m(j)-1).$$

Order these functions for each $j = 1,\dots,p$, according to descending powers of $t$, so as to obtain a vector $f(t)$, with components $f_1(t),\dots,f_n(t)$. Then the general solution of (2) is

$$x(t) = Rf(t), \tag{3}$$

where $R$ is a constant matrix. Such a solution of (2) is a solution of $x' = Ax$ if

$$Rf' = ARf. \tag{4}$$

**2. Solution with prescribed initial value.** If we wish to compute the particular solution with prescribed initial value $x(0) = c$, it is not necessary to follow the ordinary practice of computing a Jordan base and expand $c$ in that base, because the decomposition will follow automatically.

We proceed as follows. Let $X(t)$ be the matrix with columns $x(t), x'(t),\dots,x^{(n-1)}(t)$, i.e.,

$$X(t) = \mathrm{mat}(x(t),\dots,x^{(n-1)}(t)). \quad \text{(Wronskian)}$$

For a solution $x$ of $x' = Ax$ this matrix becomes

$$G(x(t)) = \mathrm{mat}(x(t),\dots,A^{n-1}x(t)).$$

Let $F(t)$ be the Wronskian of $f$, i.e.,

$$F(t) = \mathrm{mat}(f(t),\dots,f^{(n-1)}(t)).$$

From (3) and $x' = Ax$ follow the equations

$$x(t) = Rf(t),\dots,A^{n-1}x(t) = Rf^{(n-1)}(t),$$

which condense into

$$G(x(t)) = RF(t). \tag{5}$$

For $t = 0$ we have with $x(0) = c$

$$G(c) = RF(0),$$

and because $F(0)$ is nonsingular, the solution $R(c)$ for $R$ is

$$R(c) = G(c)F(0)^{-1}. \tag{6}$$

The solution $x(t)$ of $x' = Ax$ with initial value $x(0) = c$ is

$$x(t) = R(c)f(t). \tag{7}$$

**3. The matrix *F*(0).** The $m(j)$ rows in $F(0)$, corresponding to the $m(j)$-fold root $s(j)$, are

$$\begin{array}{c} f_{j,m(j)-1}(0),\dots,f_{j,m(j)-1}^{(n-1)}(0), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f_{j,0}(0),\dots,f_{j,0}^{(n-1)}(0). \end{array}$$

Since $f_{j,i}(t) = (t/i)f_{j,i-1}(t)$ for $i > 0$, we immediately find recursion relations for the computation of $f_{j,i}^{(k)}(0)$. These relations are

$$f_{j,i}^{(k)}(0) = (k/i)f_{j,i-1}^{(k-1)}(0) \qquad \text{for } i > 0,\ k > 0, \tag{8a}$$

together with

$$f_{j,i}^{(0)}(0) = 0 \qquad \text{for } i > 0, \tag{8b}$$

and

$$f_{j,0}^{(k)}(0) = s(j)^k. \tag{8c}$$

The result is the following table for the composition of $F(0)$:

$$
\begin{array}{cccccccc}
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\
0 & 0 & 0 & 0 & 1 & 5s & 15s^2 & \cdots \\
0 & 0 & 0 & 1 & 4s & 10s^2 & 20s^3 & \cdots \\
0 & 0 & 1 & 3s & 6s^2 & 10s^3 & 15s^4 & \cdots \\
0 & 1 & 2s & 3s^2 & 4s^3 & 5s^4 & 6s^5 & \cdots \\
1 & s & s^2 & s^3 & s^4 & s^5 & s^6 & \cdots
\end{array}
$$

For an $m(j)$-fold root $s(j)$ we have to use the $m(j)$ rows from the bottom, substituting $s(j)$ for $s$. The table is easily memorized by noticing the occurrence of Pascal's triangle.

**4. Comments.** The advantage of the procedure is apparent from the representations (6) and (7). Note that the solution is directly expressed in its initial value. For any $c$ the matrix $R(c)$ is easily computed, either directly or using the following formula.

If $e_i$ are the natural base vectors, i.e., $(e_i)_j = \delta_{ij}$, then we have

$$R(c) = \sum_{i=1}^{n} c_i R(e_i). \tag{9}$$

The matrices $R(e_i)$ have some interest of their own, as we shall see in the next section.

The matrix $F(0)$ depends only on (1) (and the enumeration of the roots $s(j)$) and serves for all particular solutions. Since it is a numerical matrix, it is better suited for computations than, e.g., the fundamental matrix. Note that the computation, which requires one matrix inversion and a number of matrix multiplications, does not require the construction of a Jordan base. In fact, the representation contains that part of a Jordan base that is used in the computation of $x(t)$ from its initial value $c$, and it comes naturally.

The following simple examples illustrate the main steps in the procedure:

1. Let

$$A = \begin{bmatrix} 3 & 0 & -1 \\ 2 & 2 & -1 \\ 2 & 0 & 0 \end{bmatrix},$$

then $s(1) = 1$ and $s(2) = 2$ (double).

$$F(0) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 4 \\ 1 & 2 & 4 \end{bmatrix} \text{ and } F(0)^{-1} = \begin{bmatrix} 4 & 2 & -3 \\ -4 & -3 & 4 \\ 1 & 1 & -1 \end{bmatrix}.$$

If $x(0) = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$, then $Ax(0) = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$ and $A^2x(0) = \begin{bmatrix} 8 \\ 12 \\ 8 \end{bmatrix}$, so

$$R(x(0)) = \begin{bmatrix} 2 & 4 & 8 \\ 1 & 4 & 12 \\ 2 & 4 & 8 \end{bmatrix} \begin{bmatrix} 4 & 2 & -3 \\ -4 & -3 & 4 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

Hence the solution is

$$x(t) = R(x(0))f(t) = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} te^{2t} + \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} e^{2t}.$$

The general solution can be computed either by directly computing $R(c)$ or by computing $R(e_i)$ and using (9):

$$R(e_1) = \begin{bmatrix} -1 & 0 & 2 \\ 0 & 2 & 0 \\ -2 & 0 & 2 \end{bmatrix}, \quad R(e_2) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad R(e_3) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & -1 & 0 \\ 2 & 0 & -1 \end{bmatrix}.$$

So,

$$x_c(t) = R(c)f(t) = \begin{bmatrix} (-c_1 + c_3)e^t + (2c_1 - c_3)e^{2t} \\ (2c_1 - c_3)te^{2t} + c_2e^{2t} \\ (-2c_1 + 2c_3)e^t + (2c_1 - c_3)e^{2t} \end{bmatrix}.$$

2. Let

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix},$$

then $s(1) = 2$ (triple) and $s(2) = 3$.

$$F(0) = \begin{bmatrix} 0 & 0 & 1 & 6 \\ 0 & 1 & 4 & 12 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{bmatrix} \text{ and } F(0)^{-1} = \begin{bmatrix} 12 & 6 & 9 & -8 \\ -16 & -11 & -12 & 12 \\ 7 & 6 & 6 & -6 \\ -1 & -1 & -1 & 1 \end{bmatrix}.$$

Since $A$ has many zeros, in fact is a Jordan matrix, direct computation of $R(c) = G(c)F(0)^{-1}$ is feasible. This results in

$$R(c) = \begin{bmatrix} 0 & 0 & c_1 & 0 \\ 0 & c_3 & c_2 & 0 \\ 0 & 0 & c_3 & 0 \\ 0 & 0 & 0 & c_4 \end{bmatrix},$$

from which the general solution is obtained by $x_c(t) = R(c)f(t)$.

**5. The fundamental matrix.** The representation

$$x(t) = R(x(0))f(t)$$

enables us to compute the fundamental matrix $\Phi(t)$ of the system $x' = Ax$ with very little effort. The solution $x(t)$ of $x' = Ax$ with initial value $x(0)$ is given by

$$x(t) = \Phi(t)x(0).$$

So, if we specify $x(0)$ to $e_i$, we have

$$\Phi(t)e_i = R(e_i)f(t),$$

from which we have the representation

$$\Phi(t) = \operatorname{mat}(R(e_1)f(t), \dots, R(e_n)f(t)). \tag{10}$$

Note in passing that (9) combined with (10) gives the general conversion formula

$$R(v(t))f(t) = \Phi(t)v(t). \tag{11}$$

Application of (10) to the examples gives

1. $$\Phi(t) = \begin{bmatrix} 2e^{2t} - e^t & 0 & e^t - e^{2t} \\ 2te^{2t} & e^{2t} & -te^{2t} \\ 2e^{2t} - 2e^t & 0 & 2e^t - e^{2t} \end{bmatrix},$$

2. $$\Phi(t) = \begin{bmatrix} e^{2t} & 0 & 0 & 0 \\ 0 & e^{2t} & te^{2t} & 0 \\ 0 & 0 & e^{2t} & 0 \\ 0 & 0 & 0 & e^{3t} \end{bmatrix}.$$

Formula (10) for the fundamental matrix seems to be highly preferable to the commonly used series representation, and presents a definite computational advantage in practical applications, even if one truncates the series by means of Cayley-Hamilton's theorem.

**6. Solving the system $x' = Ax + u$.** In order to solve the inhomogeneous system, we just apply variation of parameters to the general solution of the homogeneous system $x' = Ax$. So in the solution $x(t) = R(c)f(t)$ we replace $c$ by the vector $v(t)$, which we determine such that the resulting $x(t)$ is a solution to the inhomogeneous system. Then we have for $v$ the equation

$$R(v')f + R(v)f' = AR(v)f + u.$$

By linearity we have $R(v)f' = AR(v)f$, so

$$R(v')f = u.$$

Application of the conversion formula (11) gives

$$\Phi(t)v'(t) = u(t),$$

and, because $\Phi$ is nonsingular,

$$v'(t) = \Phi(t)^{-1}u(t).$$

Integrated,

$$v(t) = \int_0^t \Phi(\tau)^{-1}u(\tau)\,d\tau + c.$$

Substituting in $x(t) = R(v)f(t)$ and once more applying the conversion formula (11), we obtain the well-known formula for the solution $x_c(t)$ of $x' = Ax + u$ with $x(0) = c$:

$$x_c(t) = \Phi(t)\int_0^t \Phi(\tau)^{-1}u(\tau)\,d\tau + \Phi(t)c. \tag{12}$$

The essential point here is to compute $\Phi(t)$ according to (10).

**7. Extension.** Evidently the same procedure as described in the preceding sections can be applied to the general $n$-dimensional system of order $m$

$$x^{(m)} + A_1x^{(m-1)} + \cdots + A_mx = 0,$$

($A_1, \ldots, A_m$ constant square matrices of order $n$) after conversion into a first order $mn$-dimensional system by

$$z_k = x^{(k-1)} \qquad (k = 1, \ldots, m).$$

This results in a system

$$z' = Pz$$

of dimension $mn$.

The characteristic equation of this system is

$$\det(sI - P) = \pm\det\left(s^mI + s^{m-1}A_1 + \cdots + A_m\right) = 0.$$

This is an equation of degree $mn$ in $s$:

$$s^{mn} + a_1s^{mn-1} + \cdots + a_{mn} = 0.$$

A particular solution of the original system is specified by $x(0), x'(0), \ldots, x^{(m-1)}(0)$ and from these values the matrix

$$G(z(0)) = \operatorname{mat}\left(z(0), Pz(0), \ldots, P^{mn-1}z(0)\right)$$

can be computed.

**8. The Jordan normal form.** The role of the matrices $R(e_i)$ is not limited to the use in the composition of the general solution and the computation of the fundamental matrix. They also yield a Jordan base and a Jordan normal form. Although this is of no importance for the solution of the system, let us see how this can be done. If we write

$$Rf = \sum_{j=1}^{p} \sum_{k=0}^{m(j)-1} b_{j,k} f_{j,k},$$

then from the relation $Rf' = ARf$ we obtain the following known conditions for $b_{j,k}$

(13a) $$Ab_{j,m(j)-1} = s(j)\, b_{j,m(j)-1},$$

and for $k = 0, \ldots, m(j) - 2$:

(13b) $$Ab_{j,k} = s(j)\, b_{j,k} + b_{j,k+1}.$$

Equations (13a) and (13b) define $b_{j,k}$ as a so-called Jordan base. Both systems may have more than one solution, because a multiple root may be looked upon as a set of coinciding roots with different multiplicities, e.g., in Example 2 the threefold root 2 may be considered as two coinciding roots 2 with multiplicities 1 and 2 respectively. So from (13a) and (13b) we know that the columns of $R(e_i)$ are vectors of a Jordan base. However, in one column of the set of $R(e_i)$ there may appear vectors satisfying different systems of (13a) and (13b), and the question is how to sort them out. To do this we best look how these vectors behave under application of $A$. The result is then a Jordan base, together with the corresponding Jordan normal form of $A$, possibly after some scaling of the base vectors. By way of illustration we apply this to Examples 1 and 2, where no scaling appears to be necessary.

1. Note that $R(e_1)$ is nonsingular, so its columns already form a Jordan base. If $R(e_1) = \operatorname{mat}(b_1, b_2, b_3)$, then we see $Ab_1 = b_1$, $Ab_2 = 2b_2$, $Ab_3 = 2b_3 + b_2$. So we have the following Jordan normal form

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

2. Here for no $c$ the columns of $R(c)$ form a Jordan base. $R(e_1)$ contains one non-zero vector $e_1$ and $Ae_1 = 2e_1$. The remaining $R(e_i)$ give

$$Ae_2 = 2e_2,\ Ae_3 = 2e_3 + e_2,\ Ae_4 = 3e_4.$$

Obviously so because $A$ was already given in Jordan normal form. The natural base is a Jordan base and $A$ itself is the corresponding Jordan normal form.

The inspection may sometimes be simplified by looking at $Ax_c(t)$ or $AR(c)$. What we do is in fact the singling out of the elementary divisors and in a rather easy way.

---

**146.** **MISCELLANEA**

The Learning of this People is very defective; consisting only in Morality, History, Poetry and Mathematicks; wherein they must be allowed to excel. But, the last of these is wholly applied to what may be useful in Life; to the Improvement of Agriculture and all mechanical Arts; so that among us it would be little esteemed. And as to Ideas, Entities, Abstractions and Transcendentals, I could never drive the least Conception into their Heads.

—Jonathan Swift, *Gulliver's Travels*,
(A Voyage to Brobdingnag).

One of the founders of the theory of splines and optimal approximations. (See p. 337.)

**Note added in proof:** I should like to thank David Gale, David Hoffman, and Erwin Lutwak for interesting comments and references concerning the problem posed at the end of this paper. In particular, a paper of Tudor Zamfirescu (Proc. Amer. Math. Soc., 80, 3 (1980) 455–457) studies the number of contact points of a convex curve with its circumscribed circle and shows that in the sense of Baire categories, "most" convex curves have exactly three contact points. On the other hand, an earlier paper of Peter Gruber (Math. Ann., 229 (1977) 259–266) shows that in the same sense, "most" convex curves are not $C^2$. Thus, the Baire category approach is not appropriate to the four-vertex problem. Similarly, it is easy to see that in the $C^1$-topology, the set of convex curves with three points of contact is dense in the set of $C^1$ convex curves. However, in the $C^2$-topology there is an open set of convex curves with just two points of contact. In fact, any convex curve that contacts its circumscribed circle in exactly two antipodal points, and whose curvature at those points is strictly greater than that of the circumscribed circle, has a neighborhood (in the $C^2$-topology) with the same property. Thus, the problem of finding an appropriate measure for this question is a subtle one.

### References


1. M. Barner and F. Flohr, Der Vierscheitelsatz und seine Verallgemeinerung, Der Mathematikunterricht, 4 (1958) 43–73.
2. Marcel Berger and Bernard Costiaux, Géométrie Différentielle, Armand Colin, Paris, 1972 (see p. 362).
3. Wilhelm Blaschke, Die Minimalzahl der Scheitel einer geschlossenen konvexen Kurve, Rend. Circ. Mat. Palermo, 36 (1913) 220–222.
4. ______, Kreis und Kugel, Leipzig, 1916.
5. ______, Vorlesungen über die Differentialgeometrie I, 3. Aufl. Springer-Verlag, Berlin, 1930 (see p. 31).
6. S. S. Chern, Curves and surfaces in euclidean space, MAA Studies in Math. 4: Studies in Global Geometry and Analysis, Mathematical Association of America, Washington, DC, 1967, 16–56 (see p. 23).
7. Manfredo P. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976 (see p. 37).
8. David Fog, Über den Vierscheitelsatz und seine Verallgemeinerungen, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl., Berlin, 1933, 251–254.
9. Herman Gluck, The converse to the four-vertex theorem, L'Enseignement Math., 17 (1971) 295–309.
10. W. C. Graustein, Extensions of the four-vertex theorem, Trans. Amer. Math. Soc., 41 (1937) 9–23.
11. Otto Haupt, Zur Theorie der Ordnung reeller Kurven in der Ebene bezüglich vorgegebener Kurvenscharen, Monatsh. Math. Phys., 40 (1933) 1–53.
12. S. B. Jackson, Vertices of plane curves, Bull. Amer. Math. Soc., 50 (1944) 564–578.
13. Wilhelm Klingenberg, A Course in Differential Geometry, Springer-Verlag, New York, 1978 (see p. 28).
14. A. Kneser, Bemerkungen über die Anzahl der Extrema der Krümmung auf geschlossenen Kurven und über verwandte Fragen in einer nicht-euklidischen Geometrie, Festschrift Heinrich Weber, Teubner 1912, 170–180.
15. H. Kneser, Neuer Beweis des Vierscheitelsatzes, Christian Huygens, 2 (1922/23) 315–318.
16. Detlef Laugwitz, Differential and Riemannian Geometry, Academic Press, New York, 1965 (see p. 201).
17. Hans Mohrmann, Die Minimalzahl der Scheitel einer geschlossenen konvexen Kurve, Rend. Circ. Mat. Palermo, 37 (1914) 267–268.
18. S. Mukhopadhaya, New methods in the geometry of a plane arc, Bull. Calcutta Math. Soc., 1 (1909) 31–37.
19. ______, Extended minimum-number theorems of cyclic and sextactic points on a plane convex oval, Math. Z., 33 (1931) 648–662.
20. Peter Scherk, The four-vertex theorem, Proc. First Canadian Math. Congress, Montreal, 1945. University of Toronto Press, Toronto, 1946, 97–102.
21. Dirk J. Struik, Lectures on Classical Differential Geometry, 2nd ed., Addison-Wesley, Reading, MA, 1961 (see p. 48).
22. L. Vietoris, Ein einfacher Beweis des Vierscheitelsatzes der ebenen Kurven, Arch. Math., 3 (1952) 304–306.
23. Wolfgang Vogt, Über monotongekrümmte Kurven, J. Reine Angew. Math., 144 (1914) 239–248.


---

### ANSWER TO PHOTO ON PAGE 327

Arthur Sard (1909–1980). The picture was taken in 1964.

# TREES AND POWER-SUMS

A. H. STONE
*Department of Mathematics, University of Rochester, Rochester, NY 14627*

**1. Introduction.** Everybody knows that $3^2 + 4^2 = 5^2$. Many people must have noticed that $3^3 + 4^3 + 5^3 = 6^3$, and must have been disappointed to find that $3^4 + 4^4 + 5^4 + 6^4 \neq 7^4$. Here we claim that the "right" sequel is

$$7^4 = 6^4 + 5^4 + \frac{15}{8}4^4,$$

obtaining these equalities as special cases of a general identity. The result comes, unexpectedly, from a problem about counting trees.

**2. Graphs.** We rapidly recall some basic graphtheoretic notions and terminology, referring to [**1**], for instance, for further details. A "graph" $G$ can be thought of as a subset of 3-dimensional space, consisting of a finite non-empty set $V$ of points, the "vertices", together with a finite set $E$ of simple arcs, the "edges", each edge having its two end-points at vertices (with which the edge is "incident"). Two edges can intersect only at common end-points. It should be remarked that some authors ([**1**], for instance) allow the two end-points of an edge to coincide, so that the edge becomes a loop; but this situation does not arise in the present paper, and we have excluded it for simplicity. But we do need to allow "multiple edges"; that is, several edges may join the same two vertices.

The "degree", $d(v)$, of a vertex $v$ in a graph $G$, is the number of edges of $G$ that are incident with $v$. A vertex of degree 1 is an "end-vertex", and the (unique) edge incident with it is the corresponding "end-line". It is easy to see that the removal of an end-vertex $v$ and its end-line $e$ from a connected graph—that is, the replacement of $V$ by $V \setminus \{v\}$ and $E$ by $E \setminus \{e\}$—leaves a connected graph (for paths connecting other vertices are unaffected). (See Fig. 1.)

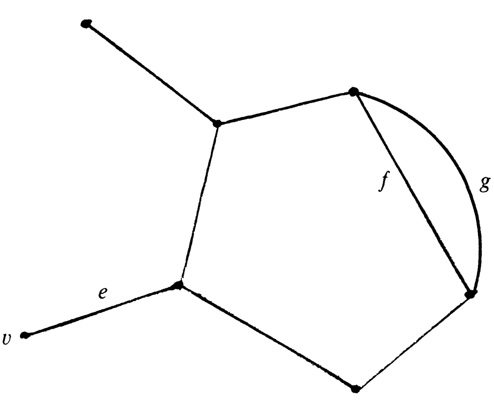



FIG. 1. The vertex $v$ is an end-vertex with end-line $e$; $f, g$ are multiple edges. There are 3 cycles (two pentagons and one digon).

---

*A. H. Stone*: I was born in London, England, and studied mathematics at Cambridge and Princeton Universities. I worked on aerodynamics at the Carnegie Institute of Washington, took up a fellowship at Trinity College, Cambridge, lectured at Manchester University, and since then have been at the University of Rochester. My family (wife, son, daughter) consists entirely of mathematicians. My research is centered on general topology; my other interests include graph theory, classical music and geometrical puzzles.

A "tree" is a connected graph that contains no simple closed paths ("cycles"). In particular, a tree can have no multiple edges. Every tree with two or more vertices has at least two end-vertices (for instance, the end-points of a simple path of greatest length); the removal of an end-vertex and its end-line from a tree will evidently leave a tree. On the other hand, the removal of an arbitrary edge $e$ from a tree $T$ will disconnect the tree (else $e$ would be part of a cycle in $T$) into exactly 2 connected components (since every vertex can be connected by a path to one or other of the end-points of $e$ without using $e$).

**3. Spanning subtrees of $K_n$.** A "complete graph" on $n$ vertices, $K_n$, is a graph with $n$ vertices each two of which are joined by exactly one edge; thus $K_n$ has exactly $\binom{n}{2}$ edges. A "spanning subtree" of a graph $G$ is a tree whose vertices are the vertices of $G$ and whose edges are among the edges of $G$. A famous theorem of Cayley [**2**] asserts that the number of different spanning subtrees of $K_n$ is $n^{n-2}$. We sketch one proof of this (due to Prüfer [**4**]), since we shall need to use the same method in a more complicated situation. The cases $n = 1, 2, 3$ are trivial, so *we assume throughout that* $n \geqslant 4$.

Suppose the vertices of $K_n$ are labelled $1, 2, \ldots, n$. For convenience of expression we do not distinguish between vertices and their labels. We associate with each spanning subtree $T$ of $K_n$ its "Prüfer symbol" $P(T)$; this will be a sequence of $n-2$ integers $t_1, t_2, \ldots, t_{n-2}$, where each $t_i$ is one of $1, 2, \ldots, n$ (repetitions allowed). $P(T)$ is defined as follows. $T$ has an end-vertex; let $s_1$ be the least end-vertex (that is, the end-vertex of $T$ with smallest label), and let $e_1$ be the corresponding end-line of $T$. We define $t_1$ to be the (label of the) vertex at the *other* end of $e_1$. Now delete $s_1$ and $e_1$ from $T$, leaving a tree $T^1$, and repeat the process on $T^1$: discard the least end-vertex $s_2$ of $T^1$, and its end-line $e_2$, to leave a smaller tree $T^2$, and define $t_2$ to be the other end-point of $e_2$; and so on. The process is stopped when $t_{n-2}$ is defined; the remaining tree $T^{n-2}$ then has just 2 vertices, and of course just one edge, the single edge joining them. Since at each stage there are at least two end-vertices available for removal, and the one removed ($s_i$) from $T^{i-1}$ is the smallest one, the vertex $n$ is never removed; thus the vertices of $T^{n-2}$ are $t_{n-2}$ and $n$ (or, if $t_{n-2} = n$, then $n-1$ and $n$).

The symbol $P(T)$ determines the spanning tree $T$ uniquely. For we obtain $T$ ($= T^0$) from $T^1$ by noting that $s_1$ is the smallest label that does not occur among $t_1, t_2, \ldots, t_{n-2}$; and $T^0$ consists of $T^1$ together with the vertex $s_1$ and the edge $s_1t_1$. Similarly $T^{i-1}$ consists of $T^i$ together with the vertex $s_i$ and edge $s_it_i$, where $s_i$ is the smallest label not occurring among $s_1, s_2, \ldots, s_{i-1}, t_i, t_{i+1}, \ldots, t_{n-2}$. Thus we can work back from $T^{n-2}$ to $T$.

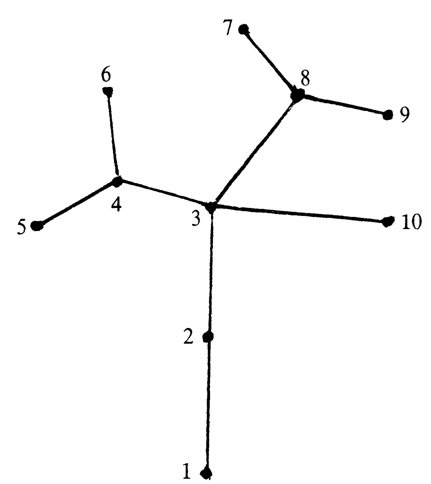



FIG. 2. Tree, with Prüfer symbol 23443883.

Moreover, an arbitrary sequence $t_1, t_2, \ldots, t_{n-2}$, where each $t_i$ is in the set $\{1, 2, \ldots, n\}$, determines a spanning subtree $T$ of $K_n$ of which it is the Prüfer symbol. For the process just described allows us to construct successively $T^{n-2}, T^{n-3}, \ldots, T^0 = T$; the point is that joining on a new end-vertex and end-line to a tree produces a tree. Hence $T^{n-2}, T^{n-3}, \ldots, T^0$ thus constructed will all be trees; and $T^0$, having $n$ vertices, will be a spanning subtree of $K_n$.

Thus the number of different spanning subtrees of $K_n$ is the number of different Prüfer symbols, $n^{n-2}$. (See Fig. 2.)

We shall need one further property of the Prüfer symbol, easily derived from the foregoing construction (see, for instance, [**1**, p. 35]): *For each vertex $v$ and spanning subtree $T$ of $K_n$, the number of times $v$ occurs in $P(T)$ is $d(v) - 1$, where $d(v)$ is the degree of $v$ in $T$.*

**4. A question and an answer.** Now consider the question: *Of the $n^{n-2}$ spanning subtrees of $K_n$, how many contain a given edge $e$?*

A first answer is as follows. Say the end-points of $e$ are $u$ and $v$. If $T$ is a spanning subtree containing $e$, the removal of $e$ splits it into two trees, say $T_1$ and $T_2$, $T_1$ having some set $V_1$ of vertices (including $u$ but not $v$) and $T_2$ having the complementary set $V_2$ of vertices (including $v$ but not $u$). For a given $V_1$, say with $r$ elements (where $1 \leqslant r \leqslant n-1$), the number of possible trees $T_1$ is $r^{r-2}$, by Cayley's theorem; and similarly the number of subtrees $T_2$ of $K_n$, having $V_2$ as vertex-set, is $(n-r)^{n-r-2}$. Any $T_1$ can be combined with any $T_2$ and the edge $e$ to produce a spanning subtree $T$ of $K_n$. Further, for a given $r$, the number of possibilities for $V_1$ is $\binom{n-2}{r-1}$. Thus the answer to the question is

$$\sum_{r=1}^{n-1} \binom{n-2}{r-1} r^{r-2} (n-r)^{n-r-2}.$$

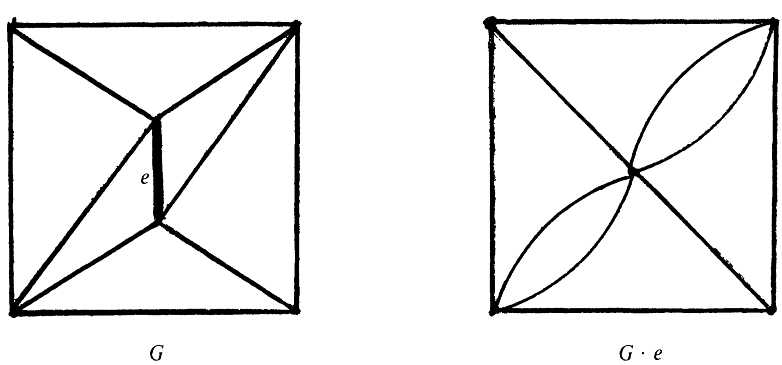



FIG. 3. A graph $(G)$ and its contraction $(G \cdot e)$.

**5. A second answer.** Another method is as follows. Form the "contracted" graph "$K_n \cdot e$" from $K_n$ by "shrinking" the edge $e$ to a point. (See Fig. 3.) That is, we discard $e$ and identify its end-points $u, v$. The resulting graph can be described as being the complete graph $K_{n-1}$ together with certain extra edges; each edge joining the vertex $u$ $(= v)$ of $K_{n-1}$ to a different vertex $w$ is replaced by two edges, corresponding to the edges $uw, vw$ of $K_n$. If $T$ is an arbitrary spanning subtree of $K_n$ containing the edge $e$, the graph $T \cdot e$ produced by this contraction is again a tree (it consists of the trees $T_1$ and $T_2$ of § 4, with the vertices $u$ and $v$ identified), and it spans $K_n \cdot e$. Conversely, every spanning subtree of $K_n \cdot e$ arises in this way from a unique spanning subtree of $K_n$ that includes the edge $e$. Thus the number of trees asked for is simply the number of spanning subtrees of $K_n \cdot e$. Now, if we ignore for the moment the doubling of the edges incident with $u$, these trees correspond to the Prüfer symbols on the $n-1$ vertices of $K_{n-1}$. However, the

doubling means that when $u$ has degree $d$ in the tree—that is, occurs $d-1$ times in the Prüfer symbol—there are $2^d$ different trees with this symbol, since in reconstructing the tree we have the choice of either of two edges on $d$ occasions. The number of these Prüfer symbols with exactly $r$ occurrences of $u$ is $\binom{n-3}{r}(n-2)^{n-r-3}$, because an arbitrary set of $r$ of the $n-3$ places in the symbol will be occupied by $u$'s, and the other $n-r-3$ places can be filled arbitrarily with the other $n-2$ vertex labels. Thus the number required is

$$\sum_{r=0}^{n-3}\binom{n-3}{r}2^{r+1}(n-2)^{n-3-r},$$

which, by the binomial theorem, is $2(2+(n-2))^{n-3}=2n^{n-3}$.

**6. The identity.** Equating the two answers, we obtain the identity (proved for $n=4,5,6,\dots$)

$$\frac{1}{2}\sum_{r=1}^{n-1}\binom{n-2}{r-1}r^{r-2}(n-r)^{n-r-2}=n^{n-3}.$$

The terms with $r=1$, $r=2$, in the sum $\Sigma$ here are, respectively, $(n-1)^{n-3}$ and $(n-2)^{n-3}$. Also, because of the symmetry property $\binom{n-2}{r-1}=\binom{n-2}{s-1}$ where $r+s=n-2$, the sum is twice the sum of its first $(n-1)/2$ terms if $n$ is odd; if $n$ is even we modify this by halving the middle term. Thus we can rewrite this identity as follows:

*If $n$ is odd, say $n=2k+1$,*

$$n^{n-3}=(n-1)^{n-3}+(n-2)^{n-3}+\cdots+\binom{n-2}{r-1}r^{r-2}(n-r)^{n-r-2}+\cdots$$
$$+\binom{n-2}{k-1}k^{k-2}(k+1)^{k-1}.$$

*If $n$ is even, say $n=2k$,*

$$n^{n-3}=(n-1)^{n-3}+(n-2)^{n-3}+\cdots+\binom{n-2}{r-1}r^{r-2}(n-r)^{n-r-2}+\cdots$$
$$+\binom{n-2}{k-2}(k-1)^{k-3}(k+1)^{k-1}+\frac{1}{2}\binom{n-2}{k-1}k^{n-4}.$$

The values $n=5,6,7,8$ give, respectively,

$$5^2=4^2+3^2,\quad 6^3=5^3+4^3+3^3,\quad 7^4=6^4+5^4+\tfrac{15}{8}4^4,$$
$$8^5=7^5+6^5+\tfrac{9}{5}5^5+\tfrac{5}{2}4^5.$$

**7. Remarks.** (1) Another (shorter but more highbrow) derivation of the answer $2n^{n-3}$, in § 5, can be obtained from the Matrix Tree Theorem (see for instance [**1**, p. 219]), which allows one to express the number of spanning subtrees of $K_n$ that do *not* contain the edge $e$ as a determinant that is not hard to evaluate.

(2) It would be interesting to have a direct elementary proof of the identity in § 6. At any rate, I do not know of one.

**References**

**1.** J. A. Bondy and U. S. R. Murty, Graph Theory with Applications, North Holland, New York, 1976.
**2.** A. Cayley, A theorem on trees, Quart. J. Pure Appl. Math., 23 (1889) 376–378.
**3.** O. Ore, Theory of Graphs, Amer. Math. Soc., Colloquium Publications 38, Providence, 1962.
**4.** H. Prüfer, Neuer Beweis eines Satzes über Permutationen, Archiv der Math. und Phys. (3) 27 (1918) 142–144.

# THE FOUR-OR-MORE VERTEX THEOREM

ROBERT OSSERMAN
*Department of Mathematics, Stanford University, Stanford, CA 94305*

The four-vertex theorem states that a smooth Jordan curve in the plane has at least four vertices. A *vertex* is a local maximum or minimum of the curvature. Thus, an ellipse has exactly four vertices, at the ends of the major and minor axes. This theorem is frequently proved, under the additional assumption that the curve is convex, in introductory differential geometry ([**2**], [**5**], [**6**], [**7**], [**13**], [**16**], [**21**]) as an early instance of a theorem requiring global rather than purely local arguments.

The four-vertex theorem (*Vierscheitelsatz*, *Théorème des quatre sommets*) has a long history, starting in 1909 with Mukhopadhaya [**18**], who stated and proved it for convex curves. There followed a succession of different proofs, generalizations, and analogies (see the References for a sample), including an interesting recent contribution due to Gluck [**9**], who proved a kind of converse. It is therefore somewhat surprising that the argument presented here seems not only to be new, but also to have a number of advantages over the usual proofs:

1. It makes immediately obvious geometrically why the result should be true.
2. It works not only for convex curves, but with only a little extra effort for arbitrary Jordan curves.
3. It is a direct proof, rather than the usual argument by contradiction. One consequence is that curves with *only* four vertices are seen to be special in certain ways; a large class of curves (even restricting to the convex case) must have six or more vertices.

The essence of the proof may be distilled in a single phrase: consider the circumscribed circle. In fact, one way to formulate the result would be the following.

THEOREM 1. *Let $\gamma$ be a smooth ($C^2$) Jordan curve in the plane. Denote by $C$ the circumscribed circle about $\gamma$. Then*
1. *$\gamma \cap C$ contains at least 2 points;*
2. *if $\gamma \cap C$ contains at least $n$ points, then $\gamma$ has at least $2n$ vertices.*

One could in fact make the second statement more precise:

THEOREM 1′. *In the notation of Theorem 1, if $R$ is the radius of $C$, and if $\gamma \cap C$ contains at least $n$ points, then either a whole arc of $\gamma$ lies on $C$, or else $\gamma$ has at least $n$ vertices where the curvature $\kappa$ satisfies $\kappa < 1/R$, and at least $n$ vertices where $\kappa \geqslant 1/R$.*

We shall discuss at the end of this paper the question of the expected number of points on $\gamma \cap C$. Note that an immediate corollary of Theorem 1 is that whenever $\gamma \cap C$ contains an infinite number of points (as it well may), $\gamma$ must have an infinite number of vertices.

The proof of Theorem 1 depends on three elementary and general geometric lemmas, as well as one lemma particular to the problem: Lemma 4 below.

LEMMA 1. *Let $E$ be a compact set in the plane containing at least two points. Then among all circles $C$ with the property that the closed disk bounded by $C$ includes $E$, there is a unique one of minimum radius $R > 0$.*

Robert Osserman received his Ph.D. at Harvard University under the supervision of Lars Ahlfors. Since then he has been on the faculty of Stanford University, with temporary or visiting positions at the University of Colorado, New York University, Harvard, and the University of Warwick. During 1960–61 he was Head of the Mathematics Branch of the Office of Naval Research; he was a Fulbright Lecturer at the University of Paris (Orsay) in 1965–66 and held a Guggenheim Fellowship in 1976–77. He spent the past year (1983–84) at the Mathematical Sciences Research Institute in Berkeley. Mathematically, his interests have always been geometry-related, starting with geometric function theory and Riemann surfaces, through minimal surfaces to general differential geometry, isoperimetric inequalities, and most recently, smooth ergodic theory.

DEFINITION. The circle defined in Lemma 1 is called the *circumscribed circle* about $E$.

LEMMA 2. *If $C$ is the circumscribed circle about $E$, then any arc of $C$ greater than a semicircle must intersect $E$.*

*Note*. The proof of Lemma 2, as well as the uniqueness of $C$ follow immediately from the observation that assuming the contrary, one could find a smaller circle enclosing $E$.

LEMMA 3. *Let a smooth oriented curve $\gamma$ have the same unit tangent at a point $P$ as a positively oriented circle $C$ of radius $R$. Let $\kappa$ be the curvature of $\gamma$. Then if $\kappa(P) > 1/R$, a neighborhood of $P$ on $\gamma$ lies inside $C$, while if $\kappa(P) < 1/R$, a neighborhood of $P$ on $\gamma$ lies outside $C$.*

We now derive Theorem 1 from these lemmas. Let $\gamma$ be a Jordan curve, $C$ the circumscribed circle, and $R$ the radius of $C$. The first statement in Theorem 1 follows immediately from Lemma 2. To prove the second statement, let $P_1, \dots, P_n$ be points of $\gamma \cap C$. If these points are ordered cyclically along $\gamma$, we obtain $n$ arcs $\gamma_1, \dots, \gamma_n$ of $\gamma$, each bounded by a pair of points on $\gamma \cap C$.

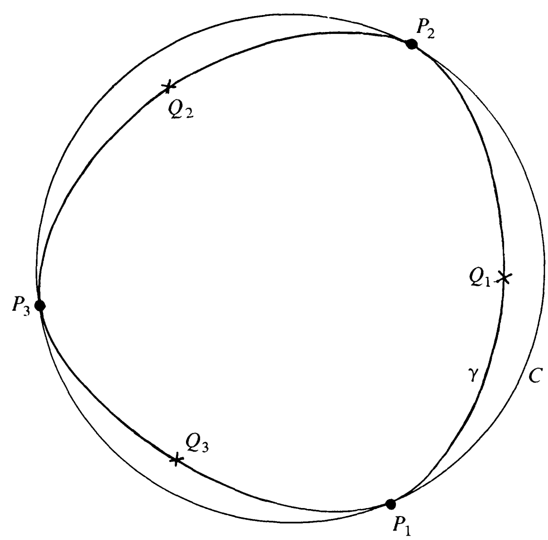



FIG. 1

*Assertion*. Each of the arcs $\gamma_i$ either lies on $C$, or else contains a point $Q_i$ such that the curvature $\kappa$ of $\gamma$ satisfies

$$\kappa(Q_i) < \frac{1}{R}. \tag{1}$$

Before proving this assertion, let us note why the theorem is an immediate consequence. First of all, we assume that $\gamma$ and $C$ both are positively oriented, so that the interior is to the left. Then at any point $P_k$ of $\gamma \cap C$, the two curves have the same orientation and $\gamma$ lies locally inside (or on) $C$. It follows from Lemma 3 that

$$\kappa(P_k) \geqslant \frac{1}{R}. \tag{2}$$

Since (2) holds at each endpoint of $\gamma_i$, it follows from (1) that $\kappa$ has a minimum at some interior point $Q_i'$ of $\gamma_i$, and that

$$\kappa(Q_i') < \frac{1}{R}. \tag{3}$$

We thus obtain $n$ vertices satisfying (3). On the other hand, each arc $\gamma_k'$ of $\gamma$ between successive $Q_i$ contains at least one point $P_k$ of $\gamma \cap C$. In view of (1) and (2), there is an interior point $P_k'$ of

$\gamma_k'$ where $\kappa$ is a maximum, and

$$\kappa(P_k') \geqslant \frac{1}{R}. \tag{4}$$

We thus get $n$ more vertices, thereby proving Theorem 1′, and hence Theorem 1. (We have ignored the possibility that one of the $\gamma_i$ lies on $C$, in which case every point of $\gamma_i$ is trivially a vertex.)

It remains to prove the Assertion above. We formulate it as a separate lemma.

LEMMA 4. *Let $\gamma$ be a positively oriented Jordan curve, $C$ the circumscribed circle and $P_1, P_2$ points of $\gamma \cap C$. Let $\gamma_1$ be the (positively oriented) arc of $\gamma$ from $P_1$ to $P_2$. Then either $\gamma_1$ coincides with the circular arc $P_1P_2$ or else there is a point $Q_1$ on $C$ satisfying* (1), *where $R$ is the radius of $C$.*

*Proof*. By Lemma 2 we may assume that the positively oriented arc of $C$ from $P_1$ to $P_2$ is included in a closed semicircle; if not, by Lemma 2, there is a point $P_2'$ between $P_1$ and $P_2$ such that the arc of $C$ from $P_1$ to $P_2'$ does lie in a (closed) semicircle, and we may apply the argument below to the subarc $\gamma_1'$ of $\gamma_1$ from $P_1$ to $P_2'$. The corresponding point $Q_1$ of $\gamma_1'$ satisfying (1) will also lie on $\gamma_1$.

For convenience of referral, assume that $C$ is centered at the origin, and that $P_1, P_2$ lie on the same vertical line in the right half-plane, with $P_2$ above $P_1$ (Fig. 2).

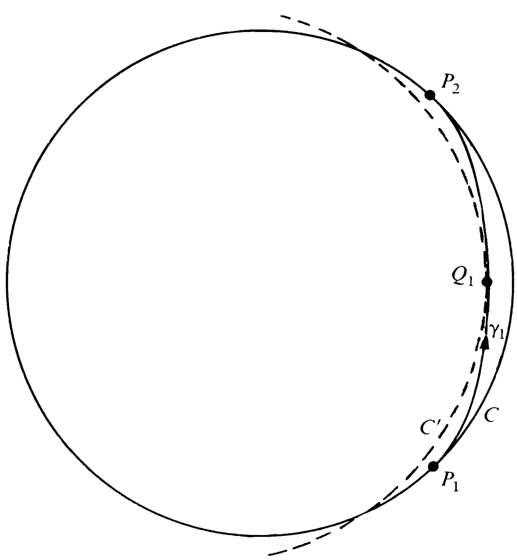



FIG. 2

There are two possibilities. Either $\gamma_1$ coincides with the circular arc $P_1P_2$, or else there is some point $Q$ on $\gamma_1$ that lies strictly inside $C$. Consider first the case where $\gamma$ is convex. If we translate the circle determined by $P_1, Q, P_2$ to the left, there will be a last moment at which it intersects $\gamma_1$. Let $C'$ be the corresponding position of the circle, and let $Q_1$ be a point of the intersection $C' \cap \gamma_1$. Since the radius $R'$ of $C'$ satisfies $R' > R$, and since $\gamma_1$ lies locally outside $C'$ at $Q_1$, it follows from Lemma 3 that

$$\kappa(Q_1) \leqslant \frac{1}{R'} < \frac{1}{R}.$$

This proves the lemma, and hence the theorem, for the case of convex curves.

Precisely the same argument holds for general Jordan curves, with one additional *caveat*: we

must use the Jordan property to guarantee that $\gamma_1$ has the same orientation as $C'$ at $Q_1$. (In fact, for non-Jordan curves that need not be the case, and the lemma, as well as the theorem, need not hold; see Fig. 3.)

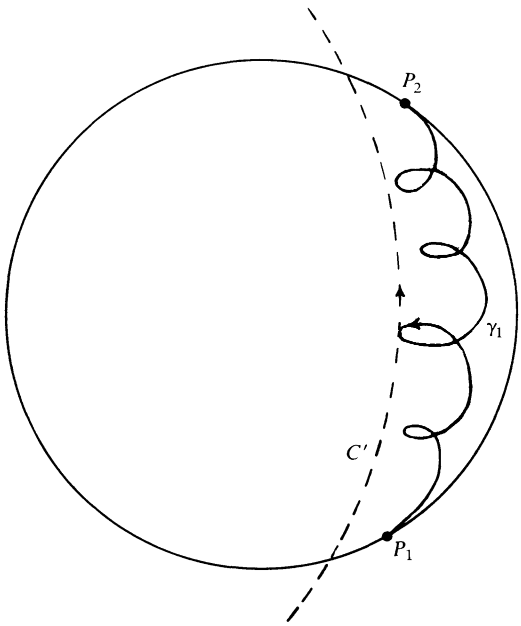



FIG. 3

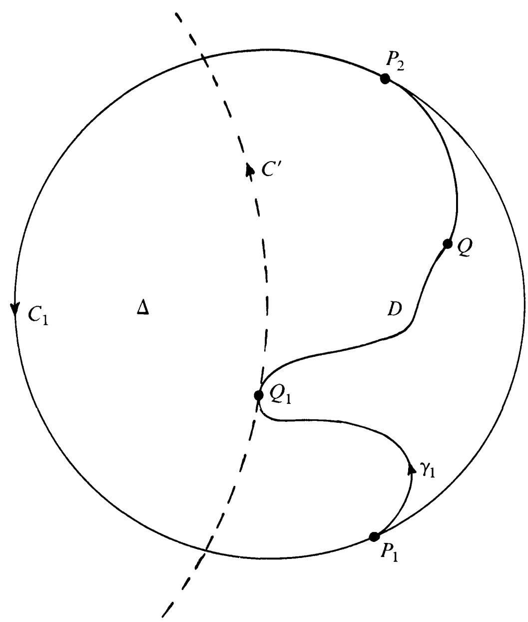



FIG. 4

Under the assumption that $\gamma_1$ has no self-intersections, the closed curve, consisting of $\gamma_1$ followed by the arc $C_1$ of $C$ going in the positive direction from $P_2$ to $P_1$, is a Jordan curve whose interior is a domain $D$ included in the interior of $C$. Note that the positive orientation induced on $\gamma_1$ as boundary of $D$ coincides with its original orientation as part of $\gamma$, since at the points $P_1$ and $P_2$, both coincide with the positive orientation of $C$. Once again, there are two cases to consider. Either $\gamma_1$ coincides with the arc of $C$ from $P_1$ to $P_2$, or else $\gamma_1$ contains a point $Q$ strictly inside $C$. In the latter case, we may choose $Q$ to the right of the vertical line through $P_1$ and $P_2$. (See also Remark 1 following the proof.) Then the circle determined by $P_1QP_2$ has radius $R' > R$. Translating this circle to the left, we again find a circle $C'$ containing a point $Q_1$ of $\gamma_1$ such that all further translates of $C'$ to the left fail to intersect $\gamma_1$. (See Fig. 4.) It follows that the interiors of $C$ and $C'$ intersect in a domain $\Delta$ that is included in $D$. Thus, both $\gamma_1$ and $C'$ have the same orientation at $Q_1$, and we may apply Lemma 2 as before to deduce

$$\kappa(Q_1) \leqslant \frac{1}{R'} < \frac{1}{R}.$$

This proves Lemma 4 and Theorem 1 for arbitrary Jordan curves.

REMARK 1. A slight modification of the argument above produces sharper quantitative results. Consider all circular arcs from $P_1$ to $P_2$ lying inside $C$. Let $C''$ be the one farthest to the left intersecting $\gamma_1$, and let $Q''$ be a point of $\gamma_1 \cap C''$. There are three cases, depending on whether $Q''$ is to the right, to the left, or on the vertical line $P_1P_2$. In the last case, the argument above shows that $\kappa(Q'') \leqslant 0$. In the other two cases, $C''$ is a proper circle of radius $R''$. If $Q''$ is to the right of the line $P_1P_2$, then again

$$\kappa(Q'') \leqslant \frac{1}{R''} < \frac{1}{R}.$$

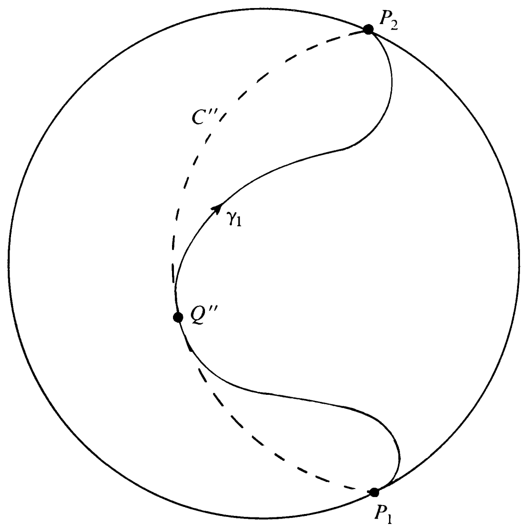



FIG. 5

If $Q''$ is the left, one has the stronger result that

$$\kappa(Q'') \leqslant -\frac{1}{R''} < 0.$$

For this last, one notes that at $Q''$, the positive orientation of $\gamma_1$ coincides with the negative orientation of $C''$ (Fig. 5).

REMARK 2. It might seem natural to carry out a basically equivalent "dual" approach, using inscribed, rather than circumscribed circles. On closer examination, however, the use of inscribed circles is considerably less straightforward. In fact, even their definition requires some care, and they are generally not unique. A paper of Jackson [**12**] contains a proof of the four-vertex theorem along those lines. He uses a proof by Erdös of the existence of specially adapted inscribed circles (p. 568). He also proves a result (Lemma 4.1) more general than our Lemma 4, making use of the Gauss-Bonnet theorem.

REMARK 3. As we said at the outset, the usual proofs of the four-vertex theorem show that the presence of fewer than four vertices would lead to a contradiction. Such proofs give no hint as to the actual number of vertices present, either on a given curve, or "in general". It follows from Theorem 1 that a curve with only four vertices must intersect its circumscribed circle in only two points. By Lemma 2, those two points must be antipodal points of the circle. Clearly, that is a fairly special property, even within the class of convex curves. Further properties that must be satisfied by curves with only four vertices have been derived by Jackson [**12**] and others. We are thus led to two questions, each of which may be considered either for the class of smooth convex curves, or more generally, for closed Jordan curves.

1. Is it more likely for a curve to have only four vertices, or to have at least six vertices?
2. Is it more likely for a curve to intersect its circumscribed circle in only two points, or in at least three points?

Intuitively, one may expect a "tripod" effect; that is, the circumscribed circle is most likely to touch the curve at *exactly* three points (see Fig. 1). We are thus led to formulate the following precise problem.

*Is there a natural measure on the space of all smooth closed curves* (*either convex or Jordan*)? *In terms of such a measure, what are the relative sizes of the sets of curves which intersect their circumscribed circles in* (a) *exactly two points*, (b) *exactly three points*, (c) *more than three points*?

**Note added in proof:** I should like to thank David Gale, David Hoffman, and Erwin Lutwak for interesting comments and references concerning the problem posed at the end of this paper. In particular, a paper of Tudor Zamfirescu (Proc. Amer. Math. Soc., 80, 3 (1980) 455–457) studies the number of contact points of a convex curve with its circumscribed circle and shows that in the sense of Baire categories, "most" convex curves have exactly three contact points. On the other hand, an earlier paper of Peter Gruber (Math. Ann., 229 (1977) 259–266) shows that in the same sense, "most" convex curves are not $C^2$. Thus, the Baire category approach is not appropriate to the four-vertex problem. Similarly, it is easy to see that in the $C^1$-topology, the set of convex curves with three points of contact is dense in the set of $C^1$ convex curves. However, in the $C^2$-topology there is an open set of convex curves with just two points of contact. In fact, any convex curve that contacts its circumscribed circle in exactly two antipodal points, and whose curvature at those points is strictly greater than that of the circumscribed circle, has a neighborhood (in the $C^2$-topology) with the same property. Thus, the problem of finding an appropriate measure for this question is a subtle one.

## References

1. M. Barner and F. Flohr, Der Vierscheitelsatz und seine Verallgemeinerung, Der Mathematikunterricht, 4 (1958) 43–73.

2. Marcel Berger and Bernard Costiaux, Géométrie Différentielle, Armand Colin, Paris, 1972 (see p. 362).

3. Wilhelm Blaschke, Die Minimalzahl der Scheitel einer geschlossenen konvexen Kurve, Rend. Circ. Mat. Palermo, 36 (1913) 220–222.

4. ______, Kreis und Kugel, Leipzig, 1916.

5. ______, Vorlesungen über die Differentialgeometrie I, 3. Aufl. Springer-Verlag, Berlin, 1930 (see p. 31).

6. S. S. Chern, Curves and surfaces in euclidean space, MAA Studies in Math. 4: Studies in Global Geometry and Analysis, Mathematical Association of America, Washington, DC, 1967, 16–56 (see p. 23).

7. Manfredo P. do Carmo, Differential Geometry of Curves and Surfaces, Prentice-Hall, Englewood Cliffs, NJ, 1976 (see p. 37).

8. David Fog, Über den Vierscheitelsatz und seine Verallgemeinerungen, Sitzungsber. Preuss. Akad. Wiss. Phys.-Math. Kl., Berlin, 1933, 251–254.

9. Herman Gluck, The converse to the four-vertex theorem, L'Enseignement Math., 17 (1971) 295–309.

10. W. C. Graustein, Extensions of the four-vertex theorem, Trans. Amer. Math. Soc., 41 (1937) 9–23.

11. Otto Haupt, Zur Theorie der Ordnung reeller Kurven in der Ebene bezüglich vorgegebener Kurvenscharen, Monatsh. Math. Phys., 40 (1933) 1–53.

12. S. B. Jackson, Vertices of plane curves, Bull. Amer. Math. Soc., 50 (1944) 564–578.

13. Wilhelm Klingenberg, A Course in Differential Geometry, Springer-Verlag, New York, 1978 (see p. 28).

14. A. Kneser, Bemerkungen über die Anzahl der Extrema der Krümmung auf geschlossenen Kurven und über verwandte Fragen in einer nicht-euklidischen Geometrie, Festschrift Heinrich Weber, Teubner 1912, 170–180.

15. H. Kneser, Neuer Beweis des Vierscheitelsatzes, Christian Huygens, 2 (1922/23) 315–318.

16. Detlef Laugwitz, Differential and Riemannian Geometry, Academic Press, New York, 1965 (see p. 201).

17. Hans Mohrmann, Die Minimalzahl der Scheitel einer geschlossenen konvexen Kurve, Rend. Circ. Mat. Palermo, 37 (1914) 267–268.

18. S. Mukhopadhaya, New methods in the geometry of a plane arc, Bull. Calcutta Math. Soc., 1 (1909) 31–37.

19. ______, Extended minimum-number theorems of cyclic and sextactic points on a plane convex oval, Math. Z., 33 (1931) 648–662.

20. Peter Scherk, The four-vertex theorem, Proc. First Canadian Math. Congress, Montreal, 1945. University of Toronto Press, Toronto, 1946, 97–102.

21. Dirk J. Struik, Lectures on Classical Differential Geometry, 2nd ed., Addison-Wesley, Reading, MA, 1961 (see p. 48).

22. L. Vietoris, Ein einfacher Beweis des Vierscheitelsatzes der ebenen Kurven, Arch. Math., 3 (1952) 304–306.

23. Wolfgang Vogt, Über monotongekrümmte Kurven, J. Reine Angew. Math., 144 (1914) 239–248.

---

### ANSWER TO PHOTO ON PAGE 327

Arthur Sard (1909–1980). The picture was taken in 1964.

# THE RESULTANT OF SEVERAL VECTORS: AN UNUSUAL INSIGHT

J. E. SZYMANSKI
*Department of Physics, University of York, Heslington, York, Y01 5DD, United Kingdom*

Of the many basic problems arising commonly in mathematics the one that I am about to consider is one of the very few that is so simple that a student with only an elementary knowledge of trigonometry can comprehend fully the steps taken at each level of its development. I hope to show, however, that even this simple treatment gives a good illustration of how an invariance to linear transformations, clearly implicit in the problem on physical grounds, manifests itself in the mathematics. It will also be seen how several points relating to computational mathematics can be introduced and how, by performing several lines of algebra, we gain new insight into an apparently trivial problem.

The problem in question is, essentially, given several vectors of known magnitudes and directions, what is the magnitude and direction of the resultant? The exact solution, of course, is

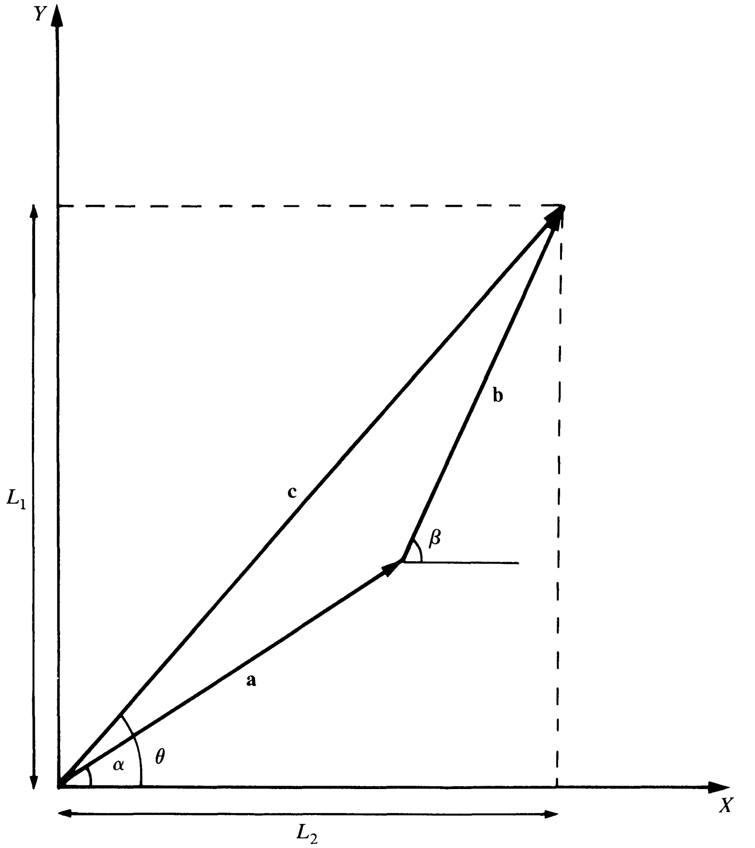



FIG. 1. The resultant vector of magnitude $||\mathbf{c}||$ and direction $\theta$ for two vectors of length $||\mathbf{a}||$ and $||\mathbf{b}||$ at angles $\alpha$ and $\beta$.

---

*J. E. Szymanski*: I received my degree in Mathematics at York in 1980, but then changed my allegiance and moved to the Physics Department where I spent my D. Phil. period working on total-energy calculations of molecules and solids under J. A. D. Matthew. I am currently concentrating on semiempirical methods of modelling the relaxation and reconstruction of solid surfaces, the mathematical basis of which involves the solution of large (of order ≈ 250) systems of coupled nonlinear equations. My leisure interests include the history of science, local history and applied horticulture (and pest control!).

trivial, but consider in more detail the special case of only two vectors. Fig. 1 illustrates this case, and immediate inspection of the diagram leads to

(1) $$\tan\theta = L_1/L_2$$

and

(2) $$\|\mathbf{c}\| = \left(L_1^2 + L_2^2\right)^{1/2},$$

where

(3) $$L_1 = \|\mathbf{a}\|\sin\alpha + \|\mathbf{b}\|\sin\beta$$

and

(4) $$L_2 = \|\mathbf{a}\|\cos\alpha + \|\mathbf{b}\|\cos\beta.$$

These expressions are exact but cumbersome, the evaluation of $\|\mathbf{c}\|$ and $\theta$ involving the calculation of four sines and cosines in addition to an arctan, a square root and ten arithmetic operations. One obvious way of improving this slightly is to rotate the coordinate axes through an angle $\alpha$, thus removing the need to evaluate $\sin\alpha$ and $\cos\alpha$ at the small cost of two more arithmetic operations. Similarly, one could divide (3) and (4) by $\|\mathbf{b}\|$, so that only the ratio $\|\mathbf{a}\|/\|\mathbf{b}\|$ enters into the problem. These are both improvements, but it is possible to do better. To show this I first present a remarkably good approximate version of (1), and in justifying this approximate expression show how the exact expression can be improved.

TABLE 1

| $\|\mathbf{a}\|$ | $\|\mathbf{b}\|$ | $\alpha$ | $\beta$ | $\theta$ | $\theta_{\text{app}}$ | $\|\mathbf{c}\|$ | $\|\mathbf{c}\|_{\text{app}}$ |
|---|---|---|---|---|---|---|---|
| 7 | 3 | 2 | 6 | 3.200 | 3.200 | 9.995 | 9.995 |
| 7 | 3 | 42 | 6 | 31.405 | 31.200 | 9.591 | 9.585 |
| 7 | 3 | 42 | 66 | 49.140 | 49.200 | 9.817 | 9.816 |
| 7 | 3 | 78 | 66 | 74.407 | 74.400 | 9.954 | 9.954 |
| 11 | 9 | 78 | 66 | 72.602 | 72.600 | 19.892 | 19.892 |
| 4 | 1 | 78 | 66 | 75.608 | 75.600 | 4.982 | 4.982 |
| 9 | 1 | 78 | 66 | 76.806 | 76.800 | 9.980 | 9.980 |
| 199 | 1 | 78 | 66 | 77.940 | 77.940 | 199.978 | 199.978 |
| 7 | 3 | 120 | 0 | 94.715 | 84.000 | 6.083 | 5.394 |
| 2 | 5 | 0 | 90 | 68.199 | 64.286 | 5.385 | 5.238 |

The approximate version of (1) in question is

(5) $$\theta_{\text{app}} = \left(\|\mathbf{a}\|\alpha + \|\mathbf{b}\|\beta\right)/\left(\|\mathbf{a}\| + \|\mathbf{b}\|\right).$$

This expression will clearly be accurate when $\alpha$ and $\beta$ (and hence $\theta$) are small angles and it is valid to replace all five of the trigonometric expressions by their small angle approximations. An examination of Table 1, however, leads to some surprises. The table lists the values of $\theta$ and $\theta_{\text{app}}$ for a range of $\|\mathbf{a}\|$, $\|\mathbf{b}\|$, $\alpha$ and $\beta$ (all angles in degrees), and the first line of the table indicates that (5) is indeed a good approximation if $\alpha$ and $\beta$ are small. The rest of the table shows, however, that the approximate expression remains remarkably accurate over a wide range of input values, the exception being the final two lines of the table. At first, it is not clear from (1) why this should be the case. There is no obvious reason why, if power series expansions of all the trigonometric functions are inserted into (1), (5) should continue to be so accurate. To see the way forward we must look again at (5), and rewrite it in the form

(6) $$\theta_{\text{app}} = \tfrac{1}{2}(\alpha + \beta) + \tfrac{1}{2}(\alpha - \beta)\left(\|\mathbf{a}\| - \|\mathbf{b}\|\right)/\left(\|\mathbf{a}\| + \|\mathbf{b}\|\right)$$

(7) $$= S + DR,$$

defining new variables

(8) $$S = \tfrac{1}{2}(\alpha + \beta),$$

(9) $$D = \tfrac{1}{2}(\alpha - \beta),$$

and

$$R = (\|\mathbf{a}\| - \|\mathbf{b}\|)/(\|\mathbf{a}\| + \|\mathbf{b}\|). \tag{10}$$

Note that, for finite $\|\mathbf{a}\|$ and $\|\mathbf{b}\|$, $R \in (-1,1)$ and $\theta_{\mathrm{app}} \in (\min(\alpha,\beta), \max(\alpha,\beta))$. Assume $\alpha \neq \beta$, else the problem is trivial.

Can (1) be written in a similar fashion? Consider the numerator

$$\|\mathbf{a}\|\sin\alpha + \|\mathbf{b}\|\sin\beta = \tfrac{1}{2}(\|\mathbf{a}\| + \|\mathbf{b}\|)(\sin\alpha + \sin\beta) + \tfrac{1}{2}(\|\mathbf{a}\| - \|\mathbf{b}\|)(\sin\alpha - \sin\beta) \tag{11}$$

$$= (\|\mathbf{a}\| + \|\mathbf{b}\|)\sin S\cos D + (\|\mathbf{a}\| - \|\mathbf{b}\|)\cos S\sin D. \tag{12}$$

Likewise, the denominator is

$$\|\mathbf{a}\|\cos\alpha + \|\mathbf{b}\|\cos\beta = (\|\mathbf{a}\| + \|\mathbf{b}\|)\cos S\cos D - (\|\mathbf{a}\| - \|\mathbf{b}\|)\sin S\sin D, \tag{13}$$

and (1) becomes

$$\tan\theta = (\tan S + R\tan D)/(1 - R\tan S\tan D). \tag{14}$$

If $(\alpha - \beta) \in (-\pi,\pi)$, then $D \in (-\pi/2,\pi/2)$ and it is possible to define a unique angle $D' \in (-\pi/2,\pi/2)$, satisfying $|D'| < |D|$, such that

$$\tan D' = R\tan D. \tag{15}$$

(14) immediately becomes

$$\tan\theta = \tan(S + D'), \tag{16}$$

so that, if $\alpha$ and $\beta \in (-\pi,\pi]$, it follows that $S \in (-\pi,\pi)$ and $\theta \in (-\pi,\pi)$ and we can write immediately

$$\theta = S + D', \tag{17}$$

since $\theta - S \in (-\pi/2,\pi/2)$. This gives

$$\theta - S = \arctan(R\tan D). \tag{18}$$

We have thus reduced (1) to the stage where it exhibits the same symmetry as (7). Equation (18) still gives $\theta$ exactly in terms of $\|\mathbf{a}\|$, $\|\mathbf{b}\|$, $\alpha$ and $\beta$, but the dependence on the absolute magnitudes of the angles has been removed into the variable $S$. In fact, the number of important variables in the problem has been reduced to two. The invariance of the answer to a rotation of the coordinate axes, not immediately obvious from (1), is reflected in the fact that the deviation of $\theta$ from the arithmetic mean of the angles $\alpha$ and $\beta$ depends only on the relative magnitude of the vectors and the angle between the vectors. Similarly the invariance to a uniform, linear scaling is reflected in the fact that $R$ depends only on $\|\mathbf{a}\|/\|\mathbf{b}\|$. It is also clear that (18) is computationally much easier to handle than (1), the calculation now only requiring two non-arithmetic operations.

Having established (18), we find that the derivation of the approximate expression given in (7) is now straightforward. The standard power series expansions of tan and arctan [**1**] lead to

$$\begin{aligned}\arctan(R\tan D) = DR\big(1 + D^2(1 - R^2)/3(1 + D^2[2 - 3R^2]/5 \\ + D^4[17 - 60R^2 + 45R^4] + \cdots)\big),\end{aligned} \tag{19}$$

where it is being assumed that $|R\tan D| < 1$. This means that this expression can only be guaranteed for all $R$ if $D \in [-\pi/4,\pi/4]$. An inspection of Fig. 2, however, seems to indicate that the range of validity increases as $|R| \to 1$. This is due to the presence of the $(1 - R^2)$ factor in all the higher order terms of the expansion. The figure shows the true accuracy embodied in the approximation of retaining only the first term of the expansion. The apparently poor results indicated in Fig. 2 can be easily reconciled with the good results contained in Table 1. The second line of the table would indicate a value of $\theta_{\mathrm{app}}$ which underestimates $\theta$ by only 0.65%. This, in fact, corresponds to an exact value of $S = 24$ and underestimating $\arctan(R\tan D)$ by 2.77%. It

should be pointed out that the worst percentage discrepancy indicated in Fig. 2 corresponds to an estimate of 4.5°, where the correct result is 5.71°. Indeed, the absolute error is worse for $R = 0.5$, where, for $D = 45°$, 26.57° is estimated as 22.5°.

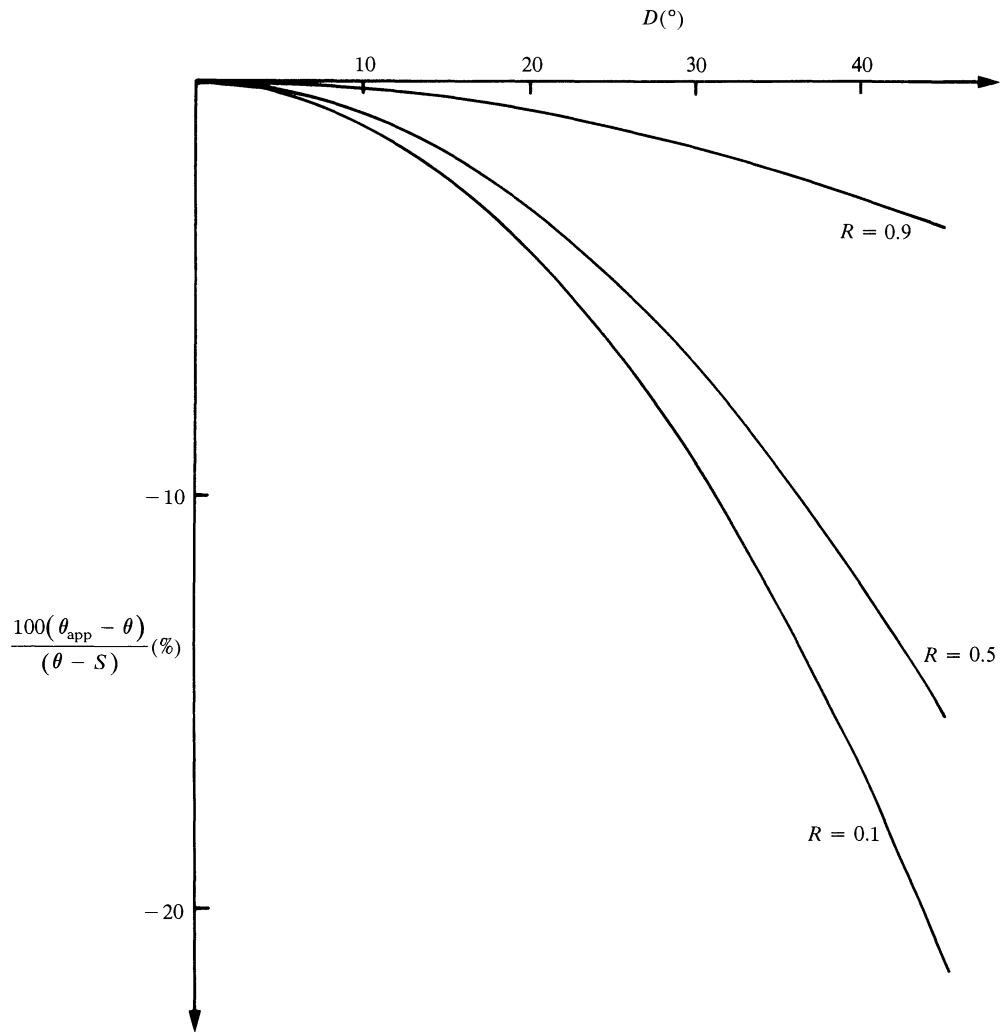



FIG. 2. The "true" accuracy of the approximation discussed in the text, defined by $100*[(\theta_{\text{app}} - S) - (\theta - S)]/(\theta - S)$.

To conclude the detailed discussion above, two points remain which need to be mentioned. The first point is that at present there is no estimate of the length of the resultant; the second point is that I have only discussed the case of two vectors. As regards the length, if (5) is to be used to calculate an approximate value of $\theta$, then $L_1$ and $L_2$, defined by (3) and (4), will never be calculated. This renders (2) useless. The only information available is a (hopefully accurate) estimate of the angle of the resultant. This is usually enough. If, instead of calculating the projections of the vectors onto the coordinate axes, we simply approximate their projections onto the resultant by

$$\|\mathbf{c}\|_{\text{app}} = (\|\mathbf{a}\| + \|\mathbf{b}\|) - \left(\|\mathbf{a}\|(\theta_{\text{app}} - \alpha)^2 + \|\mathbf{b}\|(\theta_{\text{app}} - \beta)^2\right)/2, \tag{20}$$

then an estimate for $\|\mathbf{c}\|$ is obtained which is generally very reasonable. The final two columns of

Table 1 show the correct values of $\|\mathbf{c}\|$ and those estimated by this method for a range of input parameters.

As regards the case of many vectors, it would be too much to expect a formula such as (5) to hold generally, and indeed, if a series of magnitudes and angles in the range $(-\pi, \pi)$ are chosen at random, its performance is not so good. Yet, for reasons that the analysis above makes clear, if the vectors are spread over a limited angular range of 120°, say, the results can be remarkable. The $n$-vector approximate equations are

$$\theta_{\text{app}} = \sum_{i=1}^{n} \|\mathbf{a}_i\| \alpha_i \Big/ \sum_{i=1}^{n} \|\mathbf{a}_i\| \tag{21}$$

and

$$\|\mathbf{c}\|_{\text{app}} = \sum_{i=1}^{n} \|\mathbf{a}_i\| - 0.5 \sum_{i=1}^{n} \|\mathbf{a}_i\| (\theta_{\text{app}} - \alpha_i)^2. \tag{22}$$

As an example, ten vectors with magnitudes in the range [1,10] and angles in the range [0,120] were generated at random. The vectors were

| | | |
|---|---|---|
| **V1** : | magnitude = 5, | angle = 79°, |
| **V2** : | magnitude = 2, | angle = 98°, |
| **V3** : | magnitude = 5, | angle = 88°, |
| **V4** : | magnitude = 8, | angle = 118°, |
| **V5** : | magnitude = 10, | angle = 39°, |
| **V6** : | magnitude = 3, | angle = 117°, |
| **V7** : | magnitude = 8, | angle = 21°, |
| **V8** : | magnitude = 7, | angle = 66°, |
| **V9** : | magnitude = 4, | angle = 40°, |
| **V10** : | magnitude = 6, | angle = 93°. |

TABLE 2

| | $\theta$ | $\theta_{\text{app}}$ | $\|\mathbf{c}\|$ | $\|\mathbf{c}\|_{\text{app}}$ |
|---|---|---|---|---|
| V1 + V2 | 84.398 | 84.429 | 6.922 | 6.921 |
| V1 + ⋯ + V3 | 85.909 | 85.917 | 11.916 | 11.916 |
| V1 + ⋯ + V4 | 98.718 | 98.750 | 19.171 | 19.163 |
| V1 + ⋯ + V5 | 79.089 | 78.833 | 25.707 | 25.538 |
| V1 + ⋯ + V6 | 82.846 | 82.303 | 28.134 | 27.933 |
| V1 + ⋯ + V7 | 70.381 | 70.341 | 32.680 | 32.248 |
| V1 + ⋯ + V8 | 69.609 | 69.708 | 39.663 | 39.230 |
| V1 + ⋯ + V9 | 66.986 | 67.423 | 43.186 | 42.734 |
| V1 + ⋯ + V10 | 70.087 | 70.069 | 48.649 | 48.198 |

These were added sequentially, **V1** + **V2**, **V1** + **V2** + **V3**, and so on until **V1** + **V2** + ⋯ + **V10**, and the results are shown in Table 2. The quality of the results makes the approximate formulae potentially useful for applications where many vectors are to be summed in such a fashion with absolute accuracy not necessarily demanded. Such a situation arises in the field of direct methods in crystallography [**2**], where the exact tangent formula (the $n$-vector extension of equations 1-4) forms one of the crucial equations. In this specific example the formula is used as part of an iterative cycle, so that only approximate results are required. Unfortunately, the accuracy is so unimportant in this case that a typical approach used at present to save computer time is to simply round all angles to the nearest degree and consult a table of sines and cosines held in the

computer memory. There would thus be no real effort saved by using a linearized approximate tangent formula such as (21).

Any justification of the accuracy of these formulae for several vectors would, of course, lose the clarity of the mathematics that is a feature of the two-vector case. Nevertheless, the approximations discussed here, and the insight that can be gained from analysing them, form a good example of how even the simplest of problems can lead to a useful formula and teach a great deal about a general approach to mathematics.

I would like to thank N. Anderson for pointing out the approximation and acknowledge useful discussions with H. Wright.

**References**

1. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1965, pp. 75, 81.
2. M. M. Woolfson, Phys. Bull., 34 (1983) 330.

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## THE INCLUSION-EXCLUSION PROBABILITY FORMULAS BY TAYLOR'S THEOREM

SOLOMON LEADER
*Department of Mathematics, Rutgers University, New Brunswick, NJ 08903*

Tedious combinatorial arguments in the proofs of the inclusion-exclusion probability formulas can be avoided by exploiting the fact that these formulas give the linear transformations between the coefficients of a polynomial and the coefficients of its Taylor expansion about 1. Proof of the formulas by this generating function technique requires only elementary calculus, basic properties of mathematical expectation, and the key perception that the probability sum $s_j$ in the formulas is just the expectation of $\binom{N}{j}$. Let $P$ denote probability and $E$ expectation.

THEOREM. *Given a random variable $N$ with values in $\{0,1,\dots,m\}$, let $p_k = P(N=k)$ and $r_k = P(N>k)$. Define $s_j = E\binom{N}{j}$. That is,*

$$s_j = \sum_{k=j}^{m} \binom{k}{j} p_k \text{ for } j = 0,1,\dots. \tag{1}$$

*Then*

$$p_k = \sum_{j=k}^{m} (-1)^{j-k} \binom{j}{k} s_j \text{ for } k = 0,1,\dots, \tag{2}$$

$$r_k = \sum_{j=k}^{m} (-1)^{j-k} \binom{j}{k} s_{j+1} \text{ for } k = 0,1,\dots, \tag{3}$$

*and*

$$s_{j+1} = \sum_{k=j}^{m} \binom{k}{j} r_k \text{ for } j = 0,1,\dots. \tag{4}$$

*If $N$ is the number of events among $A_1, \dots, A_m$ which occur (counting multiplicity if the events are not all distinct), then*

$$s_j = \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant m} P(A_{i_1} \cdots A_{i_j}) \text{ for } j = 1, 2, \dots . \tag{5}$$

The basic inclusion-exclusion formulas are the complementary cases $k = 0$ of (2) and (3):

$$p_0 = \sum_{j=0}^{m} (-1)^j s_j \text{ and } r_0 = \sum_{j=0}^{m} (-1)^j s_{j+1}.$$

Even for these cases the proofs are nontrivial. (See [**1**].) The case $j = 0$ of (4) gives $E(N)$ in the form $s_1 = \sum_{k=0}^{m} r_k$ that is sometimes handier than its counterpart $s_1 = \sum_{k=1}^{m} kp_k$ in (1). As systems of linear equations (2) is just the inversion of (1), and (3) the corresponding inversion of (4). The equivalence of (1) and (2) is given by the following lemma.

LEMMA 1. *For any polynomials of the form*

$$f(t) = \sum_{k=0}^{m} p_k t^k \text{ and } g(t) = \sum_{j=0}^{m} s_j (t-1)^j, \tag{6}$$

*each of the conditions* (1), (2) *is equivalent to*

$$f(t) = g(t) \text{ for all } t. \tag{7}$$

To prove Lemma 1 note that the $j$th derivative of $t^k$ is

$$k(k-1) \cdots (k-j+1) t^{k-j} = j! \binom{k}{j} t^{k-j}.$$

So for $f$ given by (6),

$$f^{(j)}(t) = j! \sum_{k=j}^{m} \binom{k}{j} p_k t^{k-j}.$$

Thus the expansion of $f$ about 1 has $f^{(j)}(1)/j! = \sum_{k=j}^{m} \binom{k}{j} p_k$ as the coefficient of $(t-1)^j$. Hence, (1) and (7) are equivalent. Similarly

$$g^{(k)}(t) = k! \sum_{j=k}^{m} \binom{j}{k} s_j (t-1)^{j-k}.$$

So the expansion of $g$ about 0 has

$$\frac{g^{(k)}(0)}{k!} = \sum_{j=k}^{m} (-1)^{j-k} \binom{j}{k} s_j$$

as the coefficient of $t^k$. Hence, (2) and (7) are equivalent.

To get (3) and (4) from (1) by Lemma 1 note that

$$p_k = r_{k-1} - r_k, \quad r_k = 1 \text{ for } k < 0, \text{ and } r_k = 0 \text{ for } k \geqslant m. \tag{8}$$

So

$$\begin{aligned}(t-1) \sum_{k=0}^{m} r_k t^k &= \sum_{k=0}^{m} r_k (t^{k+1} - t^k) = \sum_{k=1}^{m} r_{k-1} t^k - \sum_{k=0}^{m} r_k t^k \\ &= \sum_{k=1}^{m} (r_{k-1} - r_k) t^k - r_0 = \sum_{k=1}^{m} p_k t^k - (1 - p_0) \\ &= \sum_{k=0}^{m} p_k t^k - 1 = f(t) - 1\end{aligned}$$

by (6). Since $s_0 = 1$ and $s_{m+1} = 0$, (6) gives

$$g(t) - 1 = \sum_{j=1}^{m} s_j(t-1)^j = (t-1)\sum_{j=0}^{m} s_{j+1}(t-1)^j.$$

Thus, (7) is equivalent under (8) to

$$\sum_{k=0}^{m} r_k t^k = \sum_{j=0}^{m} s_{j+1}(t-1)^j \text{ for all } t. \tag{9}$$

Applying Lemma 1 with $p_k, s_j$, (1), (2) replaced respectively by $r_k, s_{j+1}$, (4), (3) we get the equivalence of (3), (4), and (9). Hence, the equivalence of (1), (2), (7), (9) implies the equivalence of (1), (2), (3), (4). To get (5) we use a simple combinatorial lemma.

LEMMA 2. *Let*

$$N = X_1 + \cdots + X_m, \tag{10}$$

*where each term $X_i$ takes its values in $\{0,1\}$. Then*

$$\binom{N}{j} = \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant m} X_{i_1} \cdots X_{i_j} \text{ for } j = 1, 2, \ldots . \tag{11}$$

$N$ is the number of terms in the sum (10) of value 1. Similarly the sum in (11) is the number of terms of value 1. Such a term in (11) has all $j$ factors chosen from the $N$ terms $X_i$ in (10) of value 1. So there are just $\binom{N}{j}$ such terms. Hence (11).

To prove (5) let $X_i$ indicate $A_i$. That is, $X_i = 1$ if $A_i$ occurs and 0 otherwise. (See [**2**].) Then (10) gives the number $N$ of events $A_1, \ldots, A_m$ which occur (counting multiplicity). Moreover, $E(X_{i_1} \cdots X_{i_j}) = P(A_{i_1} \cdots A_{i_j})$ because $X_{i_1} \cdots X_{i_j}$ indicates $A_{i_1} \cdots A_{i_j}$. So we get (5) by distributing $E$ into (11).


**References**

1. B. R. Johnson, An elementary proof of inclusion-exclusion formulas, this MONTHLY, 87 (1980) 750–751.
2. J. E. Morrill, Set theory and the indicator function, this MONTHLY, 89 (1982) 694–695.


## ON TRANSCENDENTAL LINEAR OPERATORS


A. CHARNOW AND E. CHARNOW

*Department of Mathematics, California State University-Hayward, Hayward, CA 94542*


**1.** Suppose $V$ is a vector space over a field $K$. As usual, the set of polynomials with coefficients in $K$ is denoted $K[x]$. A linear operator $T$ on $V$ will be called transcendental over $K$ if every nonzero polynomial $f$ in $K[x]$ has the property that $f(T) \neq 0$; i.e., $f(T)$ is not the zero transformation. When the dimension of $V$ is finite, a very elementary argument [**2**, p. 191] can be used to show that there are no transcendental linear operators on $V$.

If the dimension of $V$ is infinite, however, a transcendental linear operator does exist. As an example, let $B$ be a basis of $V$, let $A = \{v_i | i = 1, 2, 3, \ldots\}$ be a subset of $B$ indexed by the positive integers, and let $C$ be the complement of $A$ in $B$. Define $T$ to be the partial shift operator on $V$ satisfying $T(v_i) = v_{i+1}$ for $v_i \in A$ and $T(w) = w$ for $w \in C$. If $f$ is any nonzero polynomial in $K[x]$, then $f(T)(v_1) \neq 0$, so $f(T) \neq 0$.

If the dimension of $V$ is not only infinite but also greater than or equal to the cardinality of $K$, we have an even more dramatic change from the finite dimensional case. It is the purpose of this note to prove the following:

THEOREM. *If $V$ is an infinite dimensional vector space over a field $K$, and if the dimension of $V$ is*

**C E N T E R   S E C T I O N**
(Vol. 92, No. 5, May 1985)

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S*, L*.** "Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character. Richard P. Feynman. WW Norton, 1985, 350 pp, $16.95. [ISBN: 0-393-01921-7] A series of stories told by Nobel-laureate Feynman that reveal little science but much authentic personality--profound and funny, passionate and uninhibited. From cracking safes at Los Alamos to being judged mentally deficient by the U.S. Army, Feynman shows in these episodes the creative, quirky turn of mind that helped him become one of this century's most brilliant physicists. LAS

**General, S, L**.** A Passion To Know: 20 Profiles in Science. Ed: Allen L. Hammond. Charles Scribner's Sons, 1984, xii + 240 pp, $15.95. [ISBN: 0-684-18209-2] Fascinating insights into the minds of creative scientists, reprinted from profiles published in Science 84--the award-winning popular science magazine published by AAAS. Subjects range from astrophysics (John Wheeler) to sociology (Margaret Mead). Includes a profile of Gödel by Rudy Rucker. LAS

**General, S**, L***, P.** Mathematical People: Profiles and Interviews. Ed: Donald J. Albers, G.L. Alexanderson. Birkhauser Boston, 1985, xvi + 372 pp, $24.95. [ISBN: 3-8176-3191-7] 25 refreshingly candid portraits of contemporary mathematicians, offering personal and psychological insight into mathematical history, teaching, utility, and research. A marvelous book from which both mathematicians and non-mathematicians can gain insight into the creative process of mathematics. Many (but not all) of these interviews appeared first in The Two-Year College Mathematics Journal. LAS

**General, S*, L**, P.** Three Degrees Above Zero. Jeremy Bernstein. Charles Scribner's, 1984, xiii + 241 pp, $17.95. [ISBN: 0-684-18170-3] A window on the creative genius of Bell Labs through a series of lively vignettes of the people who work there. Begins with a profile of Ron Graham, opening Part I on computer research; moves through transistors and communication to Wilson and Penzias on black body radiation. Originally serialized in The New Yorker, this volume represents the best of its genre--popular writing about the mathematical and physical sciences. LAS

**General, S*(13), L*.** Numberplay. Gyles Brandreth. Rawson Assoc, 1984, 191 pp, $8.95 (P); $14.95. [ISBN: 0-89256-267-6] The author is a British journalist and broadcaster. He says that "The joy of numbers is that they are full of surprises," and sets out to prove it in Numberplay, a delightful confection of games to play, numbers in a variety of manifestations, stories about numbers, tricks, challenges and puzzles of all kinds. Great for browsing. JK

**Precalculus, T(13-14: 1).** Trigonometry with Applications. Dale Ewen, Lynn R. Akers. Addison-Wesley, 1984, xiii + 441 pp, $24.95. [ISBN: 0-201-11312-0] Contains computer and calculator examples. Written in a logical, straightforward fashion. Applications are mainly of typical "word-problem" variety. Has a few real-life problems. Needs some color. TR

**History, P*, L.** Ibn al-Haytham's Completion of the Conics. J.P. Hogendijk. Sources in the History of Math. & Physical Sci., V. 7. Springer-Verlag, 1985, x + 417 pp, $98. [ISBN: 0-387-96013-9] The entire Arabic text, with English translation, of Al-Hasan Ibn al-Haytham's Treatise on the Completion of the Conics, an attempted reconstruction of the lost Book VIII of Apollonius's Conics. Ibn al-Haytham (965-1041) is immortalized for his construction-solution of the now-called "problem of Alhazen (the medieval form of Al-Hasan)," from Book V of his Optics. For the specialist in the history of mathematics and science. JK

**History, P, L**.** The History of Mathematics from Antiquity to the Present: A Selective Bibliography. Joseph W. Dauben. Biblio. of the Hist. of Sci. & Tech., V. 6. Garland, 1985, xxxix + 467 pp. [ISBN: 0-8240-9284-8] An annotated bibliography of primary and secondary sources in the history of mathematics, organized by period, by discipline, by geography, and by other useful special categories. Contains over 2000 items, each with brief commentary. Part of the Smithsonian Series of Bibliographies of the History of Science and Technology. An invaluable resource to scholars and others interested in the history of mathematics. LAS

**Combinatorics, S, P.** Combinatorics and Algebra. Ed: Curtis Greene. Contemp. Math., V. 34. AMS, 1984, x + 318 pp, $30 (P). [ISBN: 0-8218-5029-6] Proceedings of the AMS-NSF conference on combinatorics and algebra held at the University of Colorado, June 5-11, 1983. Ten of the papers (including an introductory survey) impinge on the relationship between representation theory (especially of the symmetric group), and combinatorics (the Young tableaux). Remaining papers cover a wide range of topics: hyperplane arrangements, convex polytopes, finite lattices. LCL

**Algebra, S(17-18), P.** Lecture Notes in Mathematics-1093: Lectures on Formally Real Fields. Alexander Prestel. Springer-Verlag, 1984, xi + 125 pp, $7.50 (P). [ISBN: 0-387-13885-4] These republished lectures, given at the Instituto de Matemática Pura e Aplicadà in Rio de Janeiro in 1974, serve as an introduction to the theory of real fields and its connections with valuation theory and quadratic form theory. LCL

**Algebra, T(18: 1), S, P.** Linear Algebra Over Commutative Rings. Bernard R. McDonald. Pure & Appl. Math., V. 87. Dekker, 1984, viii + 544 pp, $69.75. [ISBN: 0-8247-7122-2] Starting from a fairly solid background in ring theory, the treatment serves as an introduction to and a survey of the present general theory using commutative rings as scalars. Chapters on free modules, endomorphism rings, projective modules, and the theory of a single endomorphism, through results of Bass and Kozul. Exercises, extensive bibliography, index. JS

**Algebra, S(18), P.** Quadratic and Hermitian Forms. Ed: C.R. Riehm, I. Hambleton. Conf. Proc., V. 4. AMS, 1983, xviii + 338 pp, $36 (P). [ISBN: 0-8218-6008-9] A collection of fifteen of the survey lectures, including historical remarks by Pfister, presented at a CMS conference on quadratic and hermitian forms at McMaster University in July of 1983. List of participants and technical lectures by title. JS

**Algebra, T(18: 1), S, P.** Incline Algebra and Applications. Z.-Q. Cao, K.H. Kim, F.W. Roush. Ser. in Math. & Its Applic. Halsted Pr, 1984, xiii + 165 pp, $36.95. [ISBN: 0-470-20116-9] "An assembly of all known published and unpublished results in incline algebra" relating to Boolean and fuzzy algebra including recent work of Cao. Discussion of basic concepts is followed by chapters on topological inclines, asymptotic forms, and applications. Lists of open problems, index, references. JS

**Real Analysis, T(17-18: 2), S, P, L.** Real Analysis: Modern Techniques and Their Applications. Gerald B. Folland. Wiley, 1984, xiv + 350 pp, $34.95. [ISBN: 0-471-80958-6] After a brief prologue, the treatment is fairly self-contained, proceeding through measure and integration, point set topology, and functional analysis, touching much of the fundamental theory en route. Concludes with chapters on Fourier analysis, probability, integration on groups and manifolds. Chapter notes, references, exercises. JS

**Complex Analysis, T(17-18: 1), S, P, L.** Complex Analysis: A Functional Analysis Approach. D.H. Luecking, L.A. Rubel. Universitext. Springer-Verlag, 1984, vii + 176 pp, $16 (P). [ISBN: 0-387-90993-1] An interesting and unusual second course in functions of a complex variable. The algebra of holomorphic functions on a domain, and its dual, rather than individual holomorphic functions, are the main objects of study. Nearly all results of standard graduate course are covered. 20 chapters, each with exercises. PZ

**Complex Analysis, T(18: 1, 2), S, P.** Coherent Analytic Sheaves. Hans Grauert, Reinhold Remmert. Grund. der math. Wissenschaften, B. 265. Springer-Verlag, 1984, xviii + 249 pp, $36.50. [ISBN: 0-387-13178-7] Sheaves are algebraic objects which represent the possibility of patching together local solutions to analytic problems. Coherence involves the possibility of passing from pointwise to local properties of solutions. Here sheaf theory in complex analysis is covered from its beginnings with Leray, H. Cartan, and Serre in the early fifties to the present, from main theorems to recent developments. PZ

**Complex Analysis, S(18), P*.** Kiyoshi Oka, Collected Papers. Ed: R. Remmert. Transl: R. Narasimhan. Springer-Verlag, 1984, xiv + 223 pp, $33. [ISBN: 0-387-13240-6] Ten fundamental papers (1936-1962) and two brief notes by one of the founders of function theory of several complex variables. Commentary (in French) by H. Cartan--another pioneer of the theory--follows each paper. A guide to Oka's slightly nonstandard terminology is included. PZ

**Numerical Analysis, P.** Efficient Solutions of Elliptic Systems. Ed: Wolfgang Hackbusch. Notes on Num. Fluid Mech., V. 10. Friedr. Vieweg & Sohn (US Distr: Heyden & Son), 1984, 154 pp, $20 (P). [ISBN: 3-528-08084-1] Proceedings of a January 1984 seminar held at the University of Kiel. LAS

**Functional Analysis, S(18), P.** Amenable Locally Compact Groups. Jean-Paul Pier. Wiley, 1984, x + 418 pp, $44.95. [ISBN: 0-471-89390-0] Detailed introduction to and overview of amenable locally compact (a.l.c.) groups. Presents principal characterizations of a.l.c. groups and discusses important examples of amenable groups as well as recent research results in this area. Substantial historical notes in each section and an extensive bibliography. No problems. BH

**Analysis, P.** Elliptic Boundary Value Problems. Ed: Lev. J. Leifman. Transl: J.R. Schulenberger. AMS Transl. Ser. 2, V. 123. AMS, 1984, 268 pp, $75. [ISBN: 0-8218-3082-1] Translation of seven papers by V.G. Maz'ya, B.A. Plamenevskii, et al. LAS

**Analysis, T*(16: 1), S*, L*.** Integral Transforms and their Applications, Second Edition. B. Davies. Appl. Math. Sci., V. 25. Springer-Verlag, 1985, xv + 411 pp, $28 (P). [ISBN: 0-387-96080-5]

Presents the classical integral transforms of Fourier, Hankel, Laplace, Mellin, as well as several convolution and integro-differential operators. Numerous applications, including heat and wave equations, diffraction problems, and representations of special functions. Provides simple proofs of inversion formulae for the Fourier and Laplace transforms, and other results often avoided by undergraduate texts: good supplement to courses in differential equations. Exercises. Prerequisites: some real and complex analysis. (First Edition, TR, November 1978.) YN

**Analysis, S(18), P.** Group Representations and Special Functions. Antoni Wawrzyńczyk. Math. & Its Appl. D Reidel, 1984, xvi + 688 pp, $119. [ISBN: 90-277-1269-7] A study of special functions via harmonic analysis on homogeneous spaces. Classes of special functions give rise to certain elements of Lie group representations; algebraic properties of the groups reflect analytic properties of the classes of functions. Begins with a general introduction to Lie groups and group representations. With exercises. PZ

**Geometry, P.** Few-distance Sets. A. Blokhuis. CWI Tract, No. 7. Math Centrum, 1984, iv + 70 pp, Dfl. 10,80 (P). [ISBN: 90-6196-273-0] The concept of few-distance sets stems from considering a set of points on a sphere such that the number of distinct distances occurring between pairs of the set is minimal. An extensive theory of few-distance sets has evolved for Euclidean d-space, hyperbolic space and for more esoteric spaces. This tract, presenting results of the most recent research, shows the interplay between geometry, combinatorics, algebra, operator theory and analysis in answering the questions in the field. SS

**Geometry, S*, L*.** The Fourth Dimension: Toward a Geometry of Higher Reality. Rudy Rucker. Houghton Mifflin, 1984, xi + 228 pp, $17.95. [ISBN: 0-395-34420-4] A light-hearted but nonetheless serious exploration--a "fantastic voyage"--of the fourth dimension, blending metaphysics and fantasy with geometry and physics: from Flatland to spacetime, from the occult (spirits in the fourth dimension) to the fantastic (time travel and telepathy). The promiscuous mixing of fact and fiction may feed public confusion about pseudoscience vs. science. LAS

**Optimization,** P. Lecture Notes in Mathematics-1091: Multifunctions and Integrands. Ed: G. Salinetti. Springer-Verlag, 1984, v + 234 pp, $11 (P). [ISBN: 0-387-13882-X] Proceedings of an international conference held in Catania, Italy in June 1983 to explore recent results in the theory of multi-valued functions in the context of scientific models whose solutions are not necessarily unique. LAS

**Control Theory, T(16-18), P.** Matrices in Control Theory, Revised Edition. S. Barnett. Robert E Krieger, 1984, xiv + 192 pp, $14.50. [ISBN: 0-89874-590-X] Solutions to exercises have been deleted, errors have been corrected, new references and a few new topics have been added to the 1973 original edition(TR, April 1974). LCL

**Probability, T*(15-16: 1), L.** An Introduction to Stochastic Modeling. Howard M. Taylor, Samuel Karlin. Academic Pr, 1984, x + 399 pp, $28. [ISBN: 0-12-684880-7] Designed to bridge the gap between elementary probability theory and intermediate level courses in stochastic processes. Topics covered include Markov chains in discrete and continuous time, Poisson processes, renewal theory, branching processes and queueing systems. Contains many examples and real-life applications. Good problem sets. RSK

**Probability, S(16), P.** Lecture Notes in Mathematics-1088: Characterization of Distributions by the Method of Intensively Monotone Operators. Ashot V. Kakosyan, Leo B. Klebanov, Joseph A. Melamed. Springer-Verlag, 1984, x + 175 pp, $9.50 (P). [ISBN: 0-387-13857-9] Intensively monotone operators preserve pre-ordering; here, they act on characteristic functions (Fourier transforms of densities). Their fixed points (invariant densities) give several characterizations, e.g.: if a random variable, X, has the same distribution as weighted averages of randomly many independent random variables, each distributed as X, then X must be normal. Clear enough to be accessible to advanced undergraduates, but typographical errors follow the strong law of large numbers. YN

**Probability, P.** Lecture Notes in Mathematics-1080: Probability Theory on Vector Spaces III. Ed: D. Szynal, A. Weron. Springer-Verlag, 1984, v + 373 pp, $18 (P). [ISBN: 0-387-13388-7] Proceedings of an August 1983 conference held in Lublin, Poland. LAS

**Statistics, P*.** The Collected Works of John W. Tukey, Volume I, Time Series: 1949-1964. Ed: David R. Brillinger. Stat. & Prob. Ser. Wadsworth, 1984, lxv + 689 pp, $39.95. [ISBN: 0-534-03303-2] First volume of a projected series, divided into subject areas, covering Tukey's enormous contributions to statistics. One of two volumes dealing with time series, this work includes previously unpublished papers as well as published ones, with comments on each by the editor. RSK

**Statistics, T(13), S.** The Statistical Exorcist: Dispelling Statistics Anxiety. Myles Hollander, Frank Proschan. Popular Stat., V. 3. Dekker, 1984, xi + 247 pp, $21.50. [ISBN: 0-8247-7225-3] A unique introduction to the everyday use of statistics in decision-making, sampling, analyzing data, and estimating probabilities. The text, liberally punctuated with cartoons, quotations, and anecdotes, consists of twenty-six appropriately directed vignettes (none of which makes any use of algebraic formulas) followed by related exercises. LCL

**Computer Programming, S(15-17), P*, L.** Ada for Experienced Programmers. A. Nico Habermann, Dewayne E. Perry. Addison-Wesley, 1983, xvi + 479 pp, $20.95 (P). [ISBN: 0-201-11481-X] An introduction to Ada covering the major features of the language together with their relevance to software engineering. Most chapters present a problem statement and both Pascal and Ada solutions. A particularly

valuable book for Pascal programmers who wish to learn Ada. AO

**Computer Programming, S.** Nudges, IBM Logo Projects. Steve Tipps, et al. Holt, Rinehart & Winston, 1984, ix + 273 pp, $16.95 (P). [ISBN: 0-03-000224-9] Nudges is an excellent introduction to IBM Logo. Contains many examples, hints, etc., as well as problems to solve. Written in a clear, easy-to-read style. TR

**Software Systems, S, L, P.** A User Guide to the UNIX System. Rebecca Thomas, Jean Yates. Osborne McGraw-Hill, 1985, ix + 716 pp, $18.95 (P). [ISBN: 0-88134-109-6] A greatly expanded revision of the 1982 original edition (TR, June-July 1982), emphasizing UNIX system concepts in the context of excellent tutorials on the most commonly used UNIX commands. Several useful appendices list system references--commands, tables, literature. LAS

**Software Systems, S(15-17), P.** CLU Reference Manual. B. Liskov, et al. Springer-Verlag, 1981, 190 pp, $14.95 (P). [ISBN: 0-387-91253-3] Serves as both an introduction to the programming language CLU and a reference manual for the language. The first three printings of this book appeared as Volume 114 in the "Lecture Notes in Computer Science" series. AO

**Computer Science, T(18), S, P.** Algorithms, Software, and Hardware of Parallel Computers. Ed: J. Mikloško, V.E. Kotov. Springer-Verlag, 1984, 395 pp, $32. [ISBN: 0-387-13657-6] There has been an explosive growth in the construction of parallel computing system in the last 10 years--e.g., CRAY, ETA, and Control Data. However, there has not been a parallel growth in the understanding of how to most effectively use these computers--i.e., programming languages, techniques, and algorithms. This text contains 12 research monographs covering different areas of on-going research in algorithms for parallel and array processing systems. These areas include parallel languages, correctness of parallel programs, operating systems for parallel computers, and scheduling techniques. MS

**Applications, P.** Lecture Notes in Biomathematics-55: Modelling of Patterns in Space and Time. Ed: W. Jäger, J.D. Murray. Springer-Verlag, 1984, viii + 405 pp, $23 (P). [ISBN: 0-387-13892-7] Proceedings of an interdisciplinary workshop held in Heidelberg, July 1983, covering models of dynamic patterns such as morphogenesis, chemical oscillations, fluid patterns, cell cultures, etc. LAS

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

## Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

### Northeastern Section

The fall meeting of the Northeastern Section was held on November 16-17, 1984 at Western New England College in Springfield, Massachusetts. There were 225 registrants.

Invited Addresses:

"Role of Computers in Discrete Mathematics," by John G. Kemeny, Dartmouth College.
"Role of Discrete Mathematics in the College Curriculum," by Martha Siegel, Towson State University.
"When Mathematics Says No! The Nature of Impossibilities in Mathematics," by Philip Davis, Brown University.
"Applications of Discrete Mathematics," by Fred S. Roberts, Rutgers University.

Contributed Papers:

"Open Questions About Clique Iterations," by Reverend Bruce Hedman, University of Connecticut at Hartford.

"Hybrid and Purebred Partially Ordered Sets," by C. Edward Sandifer, Western New England College.

Panel Discussion:

"Computer Science vs. Data Processing--Two-Year and Four-Year Science Curriculum," by Joseph Menard, Community College of Rhode Island (Moderator); Nelson Rich, Monroe Community College; Ken Lane, Colby College.

Student Paper:

"A Student Perspective on the Teaching of High School Computer Courses," by Mike Mair, Colby College.

Workshops:

"Exercises in Computer--Enriched Learning," by Leo F. Kuczynski, Southern Connecticut State University.
"Teaching Mathematics Using APL," by Edward LeCuyer, Jr., Western New England College.
"Introduction to Modula-2," by Hoyt Warner, Western New England College.
"Growing Mazes With Unique Solutions," by Steven Snover, University of Hartford.

## Southern California Section

The fall meeting of the Southern California Section was held at San Diego State University on November 9-10, 1984. It was a joint meeting with the American Mathematical Society.

Invited Addresses:

"Why Don't You Just Get a Bigger Computer," by Everett Bull, Pomona College.
"Symbolic Mathematics by Computer, New Opportunities for the Mathematician," by Fred Dashiell, Inference Corporation.

Short Presentation:

"Reminiscences of an Immigrant Mathematician," by Michael Golomb, Purdue University and San Diego State University.

## The Mathematical Association of America
## The Sixty Eighth Annual Meeting of the Association
## Anaheim, California

The Sixty Eighth Annual Meeting of the Association was held from January 11 through January 13, 1985, in conjunction with meetings of the American Mathematical Society, the Association for Symbolic Logic, the Association for Women in Mathematics, and the National Association for Mathematicians. The abstracts for the Invited Addresses are as follows:

The Search for Randomness, by Persi Diaconis, Stanford University.

How many times must a deck of cards be shuffled until it is close to random? It was argued that the answer is "seven." The analysis involves work of Shannon and Reeds and joint work with David Aldous. The arguments use "new wave" tools called stopping times which pervade the conversations of young probabilists. The tools involve simple symmetry arguments and handle problems that Fourier analysis can't touch.

Algorithms, Geometry, and GL($n$,$Z$), by Helaman R.P. Ferguson, Brigham Young University.

The talk began with the classical Euclidean algorithm ($n = 2$) and viewed it geometrically and algebraically. Two perpendicular lines were visualized geometrically in the plane intersecting at the origin. These two lines lie among the points of the plane with integer coordinates. The feature of the classical Euclidean algorithm was to construct pairs of integers as points in the plane on or close to this pair of lines. The Euclidean algorithm was algebraically regarded as a process to construct matrices in GL(2,Z), the 2 x 2 matrices with integer entries and determinants $\pm$ 1. Both of the latter properties relate to the geometry of the two lines and pairs of integers. The stage was then set for a generalization of the classical Euclidean algorithm to a space of $n \geq 2$ dimensions. Two perpendicular lines were replaced by a line and its perpendicular hyperplane. This configuration was situated among the n-tuples of integers. The generalization of the Euclidean algorithm which was considered here gave an iterative construction of n-tuples of integers on or close to the hyperplane and line pair: Algebraically this translated into constructing matrices in GL(n,Z), the n x n matrices with integer entries and determinant $\pm$ 1. Then n-tuples of integers on or close to the hyperplane or lines were a column or rows of a matrix of GL(n,Z) or its inverse (respectively!!!!). A proof of the existence of such iterative algorithms was given; the tools were matrix multiplication and mathematical induction. The proof was shorter than the case n = 2 of Book X of Euclid's Elements, indeed shorter than this abstract! A few applications of such algorithms given, say, decimal representations of real numbers were discussed: construct integral linear

combinations of real numbers which are zero or small; find minimal polynomials of algebraic numbers, excude small minimal polynomials; give simultaneous independent rational approximations to sets of real numbers; factor rational numbers given logarithms of primes.

Toolkit for Nonlinear Dynamics, by John M. Guckenheimer, University of California, Santa Cruz.

This was an expository lecture about methods developed to study the geometry of solutions to nonlinear systems of differential equations.

The Many Lives of Invariant Theory, by Gian-Carlo Rota, Massachusetts Institute of Technology.

A survey of current work in invariant theory from the point of view of the nineteenth century. What are invariants for? Which of these are worth classifying? How does one compute them? Do tensors have canonical forms?

Some Diophantine Problems, by Murray M. Schacher, University of California, Los Angeles.

Some of the techniques and considerations leading up to Faltings' solution of the Mordell conjecture were outlined. Applications to the Fermat problem and other questions in number theory were discussed.

Combinatorial Set Theory and Its Applications to Topology, by Franklin D. Tall, University of Toronto.

The field of set-theoretic topology has arisen in the past fifteen years as the result of the application of the new techniques of combinatorial set theory to the old field of general topology. The most notable phenomenon is the proliferation of independence results--the usual set-theoretic axioms used by mathematicians are provably inadequate to settle many of the major problems in the field. Some of the set-theoretic techniques and topological problems were discussed. The implications of independence results (which are now appearing in other areas as well) for mathematics in general were also addressed.

Some Recent Advances in Real, Complex, and Harmonic Analysis, by Guido L. Weiss, Washington University.

Many problems in analysis can be formulated as questions about the continuity of linear operators defined on certain function spaces. Such questions can be answered effectively if one can reduce them to the study of these operators on appropriate classes of simple elements that generate the entire space. Examples of such reductions are furnished by orthonormal bases for $L^2$ or by the collection of characteristic functions in $L^1$. During the past 15 years such spanning sets of simple functions have been discovered for a large number of function spaces that arise in analysis. Using these simple "building blocks" one can explain recent advances in analysis to mathematicians who are not analysts.

Trusting Computers, by Joseph Weizenbaum, Massachusetts Institute of Technology.

## Special Sessions

### Minicourses

"The Teaching of Applied Mathematics," by W. Gilbert Strang, Massachusetts Institute of Technology.

"APL--A Functional Computer Language for Mathematicians," by Garry A. Helzer, University of Maryland.

"Teaching Problem Solving," by Alan H. Schoenfeld, University of Rochester.

Applications of Discrete Mathematics," by Fred S. Roberts, Rutgers University.

"Groups, Graphs, and Computing," by Eugene M. Luks, University of Oregon.

"PROLOG," by Frederick Hoffman, Florida Atlantic University.

"Linear Programming," by Charles E. Haff, University of Waterloo.

"Teacher In-service Problems (a COMET minicourse)," by Eugene A. Maier, Mathematics Learning Center and Portland State University.

"Constructing Placement Examinations," by Richard E. Prosl, College of William and Mary.

### Panel Discussions

"How To Give A Successful Talk To Secondary School Students," sponsored by the Committee on Secondary School Lectures.

"Training Programs For Adjunct Faculty and Teaching Assistants--An Informal Exchange," sponsored by the Committee on the Teaching of Undergraduate Mathematics and organized by Bettye Anne Case, Florida State University.

"Calculus Instruction: Crucial But Ailing," jointly sponsored by AMS and MAA and organized by Ronald G. Douglas, University of California, Berkeley, and Stephen B. Maurer, Swarthmore College.

"Panel on Discrete Mathematics," led by Martha J. Siegel, Towson State University.

"Panel Report of the Fifth International Congress on Mathematical Education," presided by Donald M. Hill, Florida Agricultural and Mechanical University.

"Panel on Computers in Mathematics Education," organized by Christopher Nevison, Colgate University.

"Panel on Project Equality," co-sponsored by MAA and the National Council of Teachers of Mathematics.

Also presented: "Challenge of the Unknown," an American Association for the Advancement of Science project; "The Electronic Spreadsheet--A Creative Program For Mathematics and Mathematicians," by Deane Arganbright, Whitworth College; a special program on computer science was sponsored by the ACM/MAA Joint Committee on Retraining for Computer Science. The film "Can You Hear the Shape of a Drum?" by Mark Kac was shown in memorium.

## Contributed Paper Sessions

Teaching Introductory Statistics: Topics, Trends and Techniques. Organizer: Ann Watkins, Pierce College.

"Developing Understanding of Statistical Concepts Through Microcomputer Simulations," by Julian Weissglass, University of California, Santa Barbara.

"Using Industrial Acceptance Sampling To Introduce Hypothesis Testing in an Introductory Statistics Course," by Benjamin Fine, University of California, Santa Barbara.

"Teaching Introductory Statistics," by Clark Kimberling, University of Evansville.

"A Public Opinion Survey Project for Introductory Statistics Students," by Gary L. Britton, University of Wisconsin-West Bend.

"A Computer Based Laboratory for Introductory Statistics," by Elliot A. Tanis, Hope College.

"Computer Graphics Simulations for Elementary Probability and Statistics," by Florence S. Gordon, New York Institute of Technology, and Sheldon P. Gordon, Adelphi University.

Making Mathematics Majors Marketable: Undergraduate Training for Nonacademic Careers. Organizer: Ann K. Stehney, Institute for Defense Analyses.

"Making Mathematics Majors Marketable for a Military Service Career," by Ernest J. Manfred, U.S. Coast Guard Academy.

"Teaching Mathematical Modeling to Undergraduates at a State School," by Patricia C. Kenschaft, Montclair State College.

"Marketable Mathematics at an Undergraduate College," by Charles F. Peltier, St. Mary's College.

"Clemson's System of Options for Mathematics Majors," by Stanley M. Lukawecki, Clemson University.

"Satellite Project Brings Problems of the Workplace into the Classroom," by Lee Badger, Weber State College.

"Mathematicians at the Air Force Flight Test Center," by Henry F. Bunch, Air Force Flight Test Center, Edwards Air Force Base.

"The Applied Mathematics Laboratory," by Martha J. Siegel, Towson State University.

Strategies, Tactics, and Techniques in Teaching Lower Division and Remedial Courses. Organizer: Ann D. Holley, San Diego Mesa College.

"Visualizing Functions Without Graphs," by Martin E. Flashman, Humboldt State University.

"Using Elementary Graph Theory to Teach Non-mathematics Majors to Apply Mathematics in Problem Solving," by Sister Helen Christensen, Loyola College.

"Math As They See It: Unravelling Our Students' Thought Processes," by Deborah Hughes Hallett, Harvard University.

"Teaching as Though Students Mattered," by Alvin M. White, Harvey Mudd College.

"Teaching Estimation in Mathematics Courses for General Education," by James R. Smart, San Jose State University.

"A Direct Approach to Integration," by J.C. Smith, Virginia Polytechnic Institute and State University.

"Techniques to Foster the Development of Problem Solving Skills in Lower Division Courses," by Ann M. Chisko, University of Cinncinnati, Raymond Walters College.

"Word Problems a la Polya--Does it Work?" by Dan Kalman, Augustana College.

"The Mastery Oriented, Self-paced Instructional System for Pre-calculus Mathematics at Colorado State University," by Kenneth F. Klopfenstein, Colorado State University, Fort Collins.

"First Year Follow-up of a Modular Competency-based System for Teaching Remedial Algebra," by E. James Peake, Iowa State University.

"Students Helping Students," by Betty Mayfield, Hood College.

"Let the Computer Help," by Erik A. Schreiner, Western Michigan University.

<u>Teacher Training and Retraining</u>. Organizer: Calvin T. Long, Washington State University.

"Maximizing the Effectiveness of Teacher Retraining and Retention Efforts," Eunice Krinsky, California State University, Dominguez Hills.

"Mathematics Through Computers: An Inservice Program Designed to Integrate Computers into the Teaching and Learning of Mathematical Problem Solving," by Barbara Pence, San Jose State University.

"A Microcomputer Approach to Teaching Mathematical Abstraction to Teachers," by L.W. Linkous, University of Alabama.

"Student Teacher Correspondence: A Transformation," by Trudy Cunningham, Bucknell University.

<u>Does Research in Mathematics Learning at the College Level Exist</u>? Organizer: James J. Kaput, Southeastern Massachusetts University.

"An Overview of New Directions in Research on the Cognition of Learning and Using Mathematics," by James J. Kaput, Southeastern Massachusetts University.

"Representation of Mathematical Relationships in Four Countries: A Study of College Students' Mathematical Fluency," by Jack Lochhead, University of Massachusetts.

"Calculus Students' Utilization of Graphs as Visual Representations," by Dena Patterson, Northern Arizona University.

"College Students' Understanding of the Function Concept," by Tommy Dreyfus, San Diego State University.

"On Mathematics and Language: Why We Need to Teach the Language of the Mathematician When Teaching Mathematics," by Hadas Rin, University of California, Davis.

"Reasoning, Practice, and Success-oriented Structure in Mathematics Achievement," by Seymour W. Pustilnik, New York City Technical College.

"Heuristics and Strategies Used in Solving Equations Containing Elementary Functions," by Ronald H. Wenger, University of Delaware.

"Two Mental Orientations Manifested by Algebra and Calculus Students," by Saleh I. Assad, Erie Community College, North Campus.

"The Research Component in the Mathematics Major at Hollins College," by Caren L. Diefenderfer, Hollins College.

"Research Oversold," by S.C. Bhatnager, University of Nevada, Las Vegas.

## <u>Board and Business Meetings</u>

The Board of Governors met on January 10, 1985, and the business meeting was held on January 12, 1985. The major items of business at these meetings will be reported to the membership in FOCUS.

by (6). Since $s_0 = 1$ and $s_{m+1} = 0$, (6) gives

$$g(t) - 1 = \sum_{j=1}^{m} s_j (t-1)^j = (t-1) \sum_{j=0}^{m} s_{j+1} (t-1)^j.$$

Thus, (7) is equivalent under (8) to

$$\sum_{k=0}^{m} r_k t^k = \sum_{j=0}^{m} s_{j+1} (t-1)^j \text{ for all } t. \tag{9}$$

Applying Lemma 1 with $p_k, s_j$, (1), (2) replaced respectively by $r_k, s_{j+1}$, (4), (3) we get the equivalence of (3), (4), and (9). Hence, the equivalence of (1), (2), (7), (9) implies the equivalence of (1), (2), (3), (4). To get (5) we use a simple combinatorial lemma.

LEMMA 2. *Let*

$$N = X_1 + \cdots + X_m, \tag{10}$$

*where each term $X_i$ takes its values in $\{0,1\}$. Then*

$$\binom{N}{j} = \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant m} X_{i_1} \cdots X_{i_j} \textit{ for } j = 1, 2, \ldots . \tag{11}$$

$N$ is the number of terms in the sum (10) of value 1. Similarly the sum in (11) is the number of terms of value 1. Such a term in (11) has all $j$ factors chosen from the $N$ terms $X_i$ in (10) of value 1. So there are just $\binom{N}{j}$ such terms. Hence (11).

To prove (5) let $X_i$ indicate $A_i$. That is, $X_i = 1$ if $A_i$ occurs and 0 otherwise. (See [**2**].) Then (10) gives the number $N$ of events $A_1, \ldots, A_m$ which occur (counting multiplicity). Moreover, $E(X_{i_1} \cdots X_{i_j}) = P(A_{i_1} \cdots A_{i_j})$ because $X_{i_1} \cdots X_{i_j}$ indicates $A_{i_1} \cdots A_{i_j}$. So we get (5) by distributing $E$ into (11).

**References**

1. B. R. Johnson, An elementary proof of inclusion-exclusion formulas, this MONTHLY, 87 (1980) 750–751.
2. J. E. Morrill, Set theory and the indicator function, this MONTHLY, 89 (1982) 694–695.

## ON TRANSCENDENTAL LINEAR OPERATORS

A. CHARNOW AND E. CHARNOW
*Department of Mathematics, California State University-Hayward, Hayward, CA 94542*

**1.** Suppose $V$ is a vector space over a field $K$. As usual, the set of polynomials with coefficients in $K$ is denoted $K[x]$. A linear operator $T$ on $V$ will be called transcendental over $K$ if every nonzero polynomial $f$ in $K[x]$ has the property that $f(T) \neq 0$; i.e., $f(T)$ is not the zero transformation. When the dimension of $V$ is finite, a very elementary argument [**2**, p. 191] can be used to show that there are no transcendental linear operators on $V$.

If the dimension of $V$ is infinite, however, a transcendental linear operator does exist. As an example, let $B$ be a basis of $V$, let $A = \{v_i | i = 1, 2, 3, \ldots\}$ be a subset of $B$ indexed by the positive integers, and let $C$ be the complement of $A$ in $B$. Define $T$ to be the partial shift operator on $V$ satisfying $T(v_i) = v_{i+1}$ for $v_i \in A$ and $T(w) = w$ for $w \in C$. If $f$ is any nonzero polynomial in $K[x]$, then $f(T)(v_1) \neq 0$, so $f(T) \neq 0$.

If the dimension of $V$ is not only infinite but also greater than or equal to the cardinality of $K$, we have an even more dramatic change from the finite dimensional case. It is the purpose of this note to prove the following:

THEOREM. *If $V$ is an infinite dimensional vector space over a field $K$, and if the dimension of $V$ is*

*not less than the cardinality of $K$, then there exists a linear operator $T$ on $V$ such that $f(T)$ is invertible for every nonzero polynomial $f$ in $K[x]$.*

The key to the proof of this theorem lies in two lemmas, the first of which is used only to prove the second:

LEMMA 1. *If $K(x)$ is the field of rational functions (ratios of polynomials) over the field $K$, then as a vector space over $K$ we have* $\dim_K K(x) = \max\{\text{card } K, \aleph_0\}$.

The set of $K$-linear maps from $V$ to $V$ is denoted by $\operatorname{Hom}_K(V, V)$.

LEMMA 2. *Let $V$ be an infinite dimensional vector space over a field $K$, and assume* $\dim_K V \geqslant \text{card } K$. *If $K(x)$ is the field of rational functions over $K$, then there is a ring homomorphism* $\alpha: K(x) \to \operatorname{Hom}_K(V, V)$ *such that $\alpha(r) = r \cdot 1_V$ for all $r \in K$, where $1_V$ denotes the identity map on $V$.*

To see how the Theorem follows easily from Lemma 2, let $\alpha$ be the homomorphism described above, and set $T = \alpha(x)$. If $f$ is any nonzero polynomial in $K[x]$, then the application of $\alpha$ to $ff^{-1} = 1 = f^{-1}f$ yields

$$f(T)\alpha(f^{-1}) = 1_V = \alpha(f^{-1})f(T).$$

Thus $f(T)$ is invertible, as desired.

## 2. Proofs of the lemmas

*Proof of Lemma* 1. Let $S = \{(x-a)^{-1} | a \in K\}$. The linear independence of $S$ over $K$ is easy to verify. Choose a basis $B$ of $K(x)$ over $K$ containing $S$. Then

$$\dim_K K(x) = \text{card } B \geqslant \text{card } S = \text{card } K.$$

Because $\dim_K K(x)$ is not finite, $\dim_K K(x) \geqslant \text{card } K$ implies $\dim_K K(x) = \dim_K K(x) \text{card } K$. From the infinite dimensionality of $K(x)$ over $K$ we may also conclude that $\dim_K K(x) \text{card } K = \text{card } K(x)$ [**3**, Lemma 1, p. 245]. Therefore, it suffices to prove:

$$\text{card } K(x) = \max\{\text{card } K, \aleph_0\}. \tag{1}$$

Because a polynomial in $K[x]$ of degree $n$ is a sequence of $n+1$ elements from $K$, we have [**1**, Example 8, p. 25]

$$\text{card } K[x] = \aleph_0 \text{ card } K = \max\{\text{card } K, \aleph_0\}.$$

Any element $u \in K(x)$ can be written uniquely as $u = fg^{-1}$, where $f$ and $g$ are relatively prime polynomials with $g$ monic. The mapping $u \to (f, g)$ is an injection of $K(x)$ into $K[x] \times K[x]$. Hence

$$\text{card } K(x) \leqslant \text{card } K[x] \text{ card } K[x] = \max\{\text{card } K, \aleph_0\}.$$

Obviously $\text{card } K(x) \geqslant \max\{\text{card } K, \aleph_0\}$, so (1) is established.

*Proof of Lemma* 2. We shall first show that the scalar multiplication on $V$ can be extended in a manner which makes $V$ a vector space over $K(x)$. The hypothesis of this lemma is that $\max\{\text{card } K, \aleph_0\} \leqslant \dim_K V$. Therefore, by Lemma 1, the vector space $K(x)$ over $K$ satisfies $\dim_K K(x) \leqslant \dim_K V$, and since $\dim_K V$ is infinite, we have $\dim_K K(x) \dim_K V = \dim_K V$. Let $B$ be a basis of $K(x)$ over $K$. Choose a vector space $V'$ over $K(x)$ with $\dim_{K(x)} V' = \dim_K V$, and let $B'$ be a basis of $V'$ over $K(x)$. If $S = \{bb' | b \in B, b' \in B'\}$, then $S$ can easily be shown to be a basis of $V'$ over $K$. It follows that

$$\dim_K V' = \text{card } S = \text{card } B \text{ card } B' = \dim_K K(x) \dim_K V = \dim_K V.$$

Thus $V$ and $V'$ are $K$-isomorphic. Let $\phi$ be a $K$-isomorphism of $V$ onto $V'$. For $s \in K(x)$ and $v \in V$, define $s \circ v \in V$ to be $\phi^{-1}(s(\phi(v)))$. Clearly $r \circ v = rv$ if $r \in K$, so $\circ$ extends the original scalar multiplication on $V$. Moreover, if $s \in K(x)$, $s' \in K(x)$, and $v \in V$, then

$$(ss')\circ v = \phi^{-1}((ss')\phi(v)) = \phi^{-1}(s(s'\phi(v))) = \phi^{-1}\big(s\big(\phi\phi^{-1}(s'\phi(v))\big)\big)$$
$$= s\circ\big(\phi^{-1}(s'\phi(v))\big) = s\circ(s'\circ v).$$

The other vector space axioms are verified in a similar fashion; therefore $V$ is a vector space over $K(x)$.

For $s \in K(x)$, let $\alpha_s$ be the mapping of $V$ into $V$ defined by $\alpha_s(v) = s\circ v$ for $v \in V$. Clearly, $\alpha_s \in \mathrm{Hom}_K(V,V)$. If $\alpha\colon K(x) \to \mathrm{Hom}_K(V,V)$ is defined by $\alpha(s) = \alpha_s$ for $s \in K(x)$, then $\alpha$ is a ring homomorphism satisfying $\alpha(r) = r\cdot 1_V$ for all $r \in K$, and the proof is complete.

**References**

1. J. Goldhaber and G. Ehrlich, Algebra, Macmillan, Toronto, 1970.
2. K. Hoffman and R. Kunze, Linear Algebra, 2nd ed., Prentice-Hall, Englewood Cliffs, 1971.
3. N. Jacobson, Lectures in Abstract Algebra, Volume II-Linear Algebra, Van Nostrand, New York, 1953.

## DINI TYPE THEOREMS ON POSITIVE SERIES

PIOTR BILER
*Institute of Mathematics, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland*

Consider a convergent series $\Sigma_{n=1}^{\infty} a_n$ with positive terms $a_n > 0$. Let $r_n$ be the tail $r_n = \Sigma_{m=n}^{\infty} a_m$, and let $t$ be a positive number. The classical Dini theorem [1] states that $\Sigma_{n=1}^{\infty} a_n/r_n^t$ converges if $t < 1$ and diverges if $t \geqslant 1$. Our aim in this paper is to extend Dini's theorem by characterizing all functions $F$: $\mathbb{R}^+ \to \mathbb{R}^+$ such that $\Sigma_{n=1}^{\infty} a_n F(r_n)$ converges for every positive convergent series $\Sigma_{n=1}^{\infty} a_n$. Theorem 1 accomplishes this goal. The complementary results for divergent series are contained in Theorem 2.

THEOREM 1. *Let $F$ be a positive function on $\mathbb{R}^+$. The following conditions are equivalent*:

(1) *For any convergent positive series $\Sigma a_n$, the series $\Sigma a_n F(r_n)$ is also convergent.*

(2) *There exist a $\delta > 0$ and an increasing positive function $f$ on $(0,\delta]$ such that*

$$F(x) \leqslant (f(x) - f(y))/(x-y) \text{ for all } 0 < y < x \leqslant \delta.$$

(3) *There exist a $\delta > 0$ and a decreasing function $h$ on $(0,\delta)$ such that $h$ is integrable on $(0,\delta)$ and $h$ majorizes $F$: $F \leqslant h$.*

*Proof*. Observe that (3) implies (1) trivially:

$$\sum_{n=m}^{\infty} a_n F(r_n) \leqslant \sum_{n=m}^{\infty} (r_n - r_{n+1}) h(r_n) \leqslant \int_0^{r_m} h(x)\,dx < \infty$$

for any $m$ such that $r_m < \delta$.

Now we assume that the condition (1) holds. First we shall show the existence of a $\delta > 0$ such that $\Sigma a_n \leqslant \delta$ implies $\Sigma a_n F(r_n) \leqslant 1$. If no such $\delta$ exists, then there is a positive sequence $\langle a_{n,1}\rangle$ satisfying

$$\delta_1 = \sum_{n=1}^{\infty} a_{n,1} < 2^{-1} \text{ and } \sum_{n=1}^{n(1)} a_{n,1} F(r_{n,1}) > 1$$

for suitably large $n(1)$, where $r_{n,k} = \Sigma_{m=n}^{\infty} a_{m,k}$. Likewise there is a positive sequence $\langle a_{n,2}\rangle$ with the property

$$\delta_2 = \sum_{n=1}^{\infty} a_{n,2} < \min\left(2^{-2}, \delta_1 - \sum_{n=1}^{n(1)} a_{n,1}\right)$$

Adding up these last inequalities, we get $\sum_{n=1}^{\infty} t_n \leqslant b_1 \leqslant f(1)$ and so $\int_0^1 h < \infty$. The proof of Theorem 1 is now complete.

Below $s_n$ denotes the partial sum $s_n = \sum_{m=1}^{n} a_m$ of a series $\sum_{n=1}^{\infty} a_n$.

THEOREM 2. *Let $F$ be a positive function on $\mathbb{R}^+$. The following conditions are equivalent*:

(1′) *For any divergent positive series $\sum a_n$, the series $\sum a_n F(s_n)$ is convergent.*
(2′) *There exist a $\delta > 0$ and a decreasing function $f$ on $[\delta, \infty)$ such that*

$$F(x) \leqslant (f(y) - f(x))/(x - y) \text{ for all } \delta \leqslant y < x.$$

(3′) *There exists a decreasing function $h$, integrable on $(\delta, \infty)$, $\delta > 0$, which majorizes $F$: $F \leqslant h$.*

The idea of the proof is similar to that in Theorem 1. If (1′) holds, there exists a $\delta > 0$, perhaps large, such that the inequality $s_m \geqslant \delta$ implies $\sum_{n=m+1}^{\infty} a_n F(s_n) \leqslant 1$. The function $f$ is now defined on $[\delta, \infty)$ by $f(x) = \sup \sum_{n=m+1}^{\infty} a_n F(s_n)$, the supremum being taken over all divergent positive series $\sum a_n$ with $\sum_{n=1}^{m} a_n = x$. If (2′) holds, we put

$$g(x) = \inf_{\delta \leqslant y < x} ((f(y) - f(x))/(x - y)),$$

and then the function $h(x) = \sup_{x \leqslant t} g(t)$ is decreasing and integrable on $(\delta, \infty)$. All other details of the proof can be written as before. We leave them to the reader.

**Acknowledgement.** The author wishes to thank Professor C. Ryll-Nardzewski for suggesting a possible generalization of the Dini theorems.

**Reference**

1. Konrad Knopp, Theory and Application of Infinite Series, Blackie and Son, London, 1966.

## CLASSIFICATION OF SURFACES

C.E. BURGESS
*Department of Mathematics, University of Utah, Salt Lake City, Utah 84112*

We use "surface" synonymously with "compact connected 2-manifold without a boundary", i.e., a surface is a compact connected metric space that is locally homeomorphic with the Euclidean plane $E^2$. The purpose of this note is to prove the following classification theorem, first discovered in the early part of the twentieth century. (See comments and references in [**3**, p. 53].)

THEOREM 1. *Every surface is homeomorphic with a space obtained by removing a finite number of disjoint disks from a 2-sphere and replacing each of them with a Möbius band or a punctured torus.*

The first step of the proof is to model the given surface with a polygonal disk $D$ whose edges are identified in pairs. The existence of such a model depends on the fact that every surface can be triangulated. (See comments and references in [**3**, p. 52].) The remainder of the proof usually rests on a tedious process of cutting and pasting operations, performed on $D$, that produce another disk $D'$ so that the given surface results from identifying, in pairs, the edges of $D'$ in such a manner that all vertices of $D'$ are identified at a single point and any two identified edges are adjacent on the boundary of $D'$. (Discussions of this procedure can be found in any of the four references.)

In this note, we give an alternative proof by using induction on the number of edges of the disk $D$. For this, it will be convenient to use the concept of "a connected sum of two surfaces". (See pp. 8–10 in [**3**] and the first definition below.)

DEFINITIONS AND NOTATION. Let $M_1$ and $M_2$ be two disjoint surfaces and $D_1$ and $D_2$ be disks in $M_1$ and $M_2$, respectively. Let $M = (M_1 - \operatorname{Int} D_1) \cup (M_2 - \operatorname{Int} D_2)$, where $M_1 - \operatorname{Int} D_1$

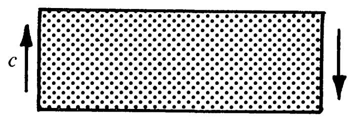



FIG. 1.1

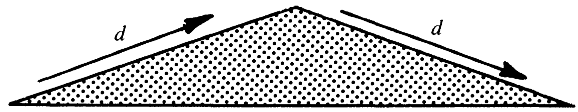



FIG. 1.2

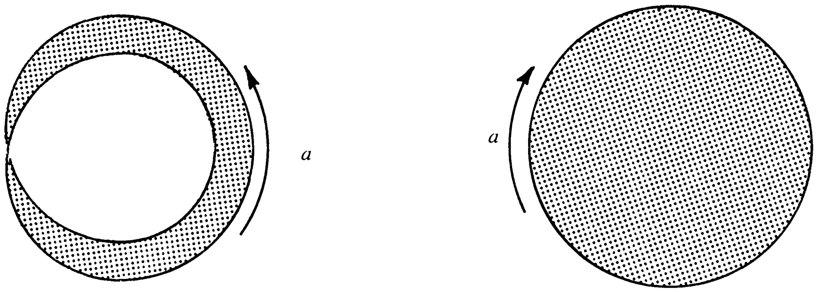



FIG. 2

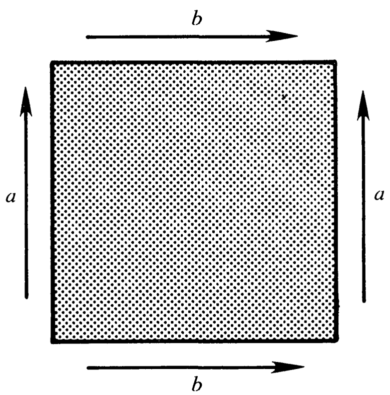



FIG. 3

and $M_2 - \operatorname{Int} D_2$ are identified (sewn together) on their boundaries, i.e., for some homeomorphism $h$ of $\operatorname{Bd} D_1$ onto $\operatorname{Bd} D_2$, $M$ is the quotient space

$$[(M_1 - \operatorname{Int} D_1) \cup (M_2 - \operatorname{Int} D_2)]/\{(x, h(x)) | x \in \operatorname{Bd} D_1\}.$$

The surface $M$ is called the connected sum of $M_1$ and $M_2$ and is denoted by $M_1 \# M_2$. (It is well known that $M_1 \# M_2$ is independent of the choice of $D_1$ and $D_2$ and that

$$M_1 \# (M_2 \# M_3) = (M_1 \# M_2) \# M_3,$$

but it is not our purpose in this note to elaborate on these properties.)

A **2-sphere**, denoted by $S$, is a space that is homeomorphic with the graph of $x^2 + y^2 + z^2 = 1$ in $E^3$.

A **Möbius band** is a space obtained by identifying, or sewing, two opposite edges of a rectangular disk as indicated in Fig. 1.1. Equivalently, a Möbius band is obtained by identifying two adjacent edges of a triangular disk as indicated in Fig. 1.2. To see that a Möbius band results in 1.2, draw a diagonal in the rectangle of 1.1, form the Möbius band by identifying the two ends as indicated, and then cut along the simple closed curve that results from identifying the end points of this diagonal.

A **projective plane**, denoted by $P$, is a space obtained by sewing a Möbius band and a disk

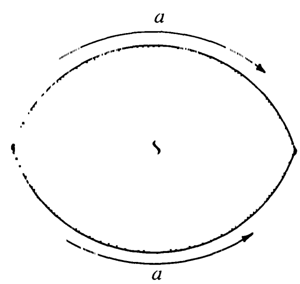



FIG. 4.1

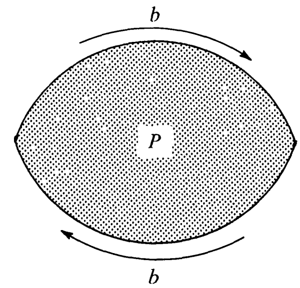



FIG. 4.2

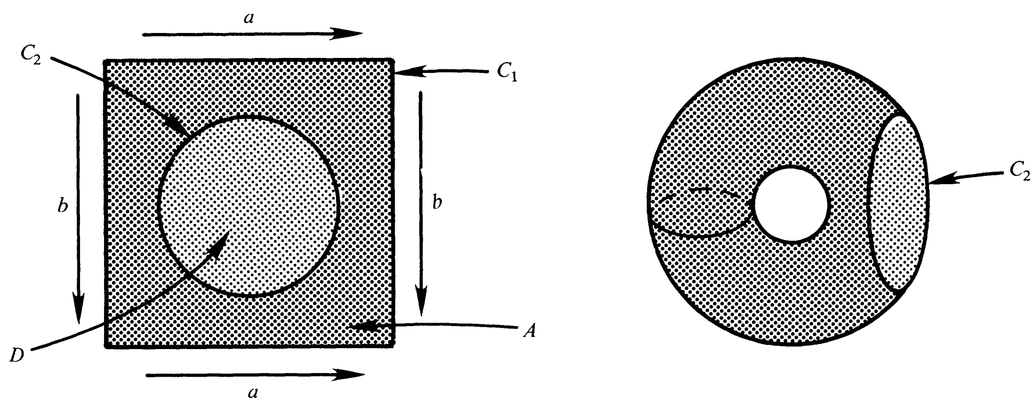



FIG. 5

together on their boundaries, as in Fig. 2. Equivalently, a projective plane is obtained with the identification indicated in Fig. 4.2.

A **torus**, denoted by $T$, is a space homeomorphic with the Cartesian product $S^1 \times S^1$, where $S^1$ denotes a circle. A torus is obtained by identifying the four edges of the square disk as indicated in Fig. 3.

Let $D(n)$ denote a disk with $n$ edges on its boundary, where $n$ is even, and let $M(n)$ denote a surface that is obtained by identifying, in pairs, the edges of $D(n)$.

Two identified edges of a disk are called a **twisted pair** if the identification involves the same direction for the two edges around the boundary of $D$. The pairs labeled with $c$ and $d$ in Fig. 1 and with $a$ in Fig. 6.1 are illustrations of twisted pairs.

Two pairs of identified edges of a disk $D$ are called **separated pairs** if the two edges in one pair separate the two in the other pair on the boundary of $D$. The pairs labeled with $a$ and $b$ in Figs. 3 and 7.1 and with $b$ and $c$ in Fig. 6.1 are illustrations of separated pairs that are nontwisted.

The second step in the proof of Theorem 1 is to present two lemmas that are used in the inductive procedure described in the fourth step.

LEMMA 1. *If $M(2)$ is a surface obtained by identifying the two edges of a disk $D(2)$, then $M(2)$ is either a sphere or a projective plane.*

There are two different identifications as indicated in Fig. 4, with the two resulting surfaces—the sphere and the projective plane—labeled $S$ and $P$.

LEMMA 2. *If $A$ is an annulus with $n$ edges ($n$ even) on one component $C_1$ of its boundary, then any space obtained by identifying these $n$ edges in pairs is homeomorphic with a punctured $M(n)$, i.e., there is a disk $D$ in $M(n)$ such that the resulting space is homeomorphic with $M(n) - \operatorname{Int} D$.*

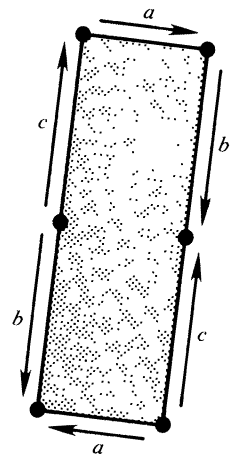



FIG. 6.1

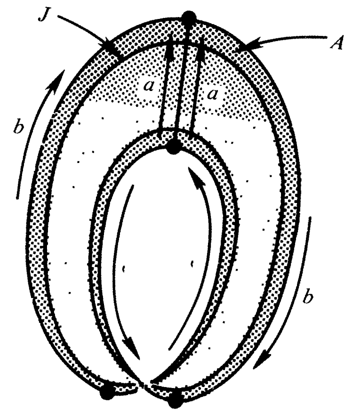



FIG. 6.2

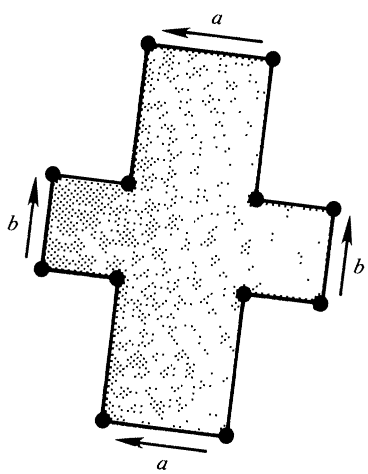



FIG. 7.1

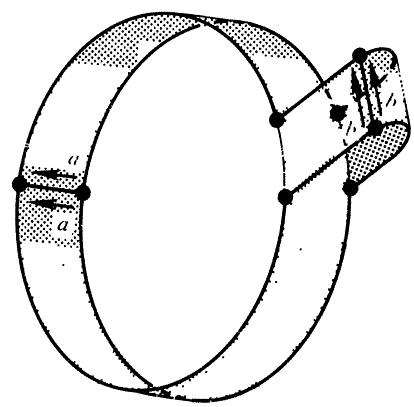



FIG. 7.2

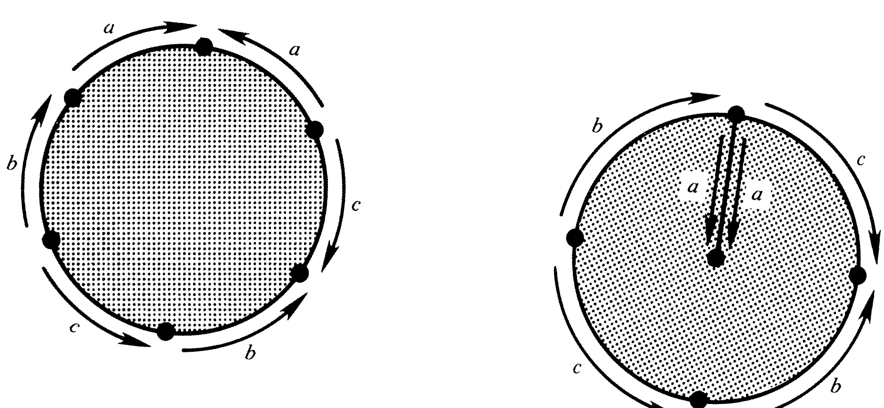



FIG. 8

This is easily seen by identifying the boundary of a disk $D$ with the other component $C_2$ of the boundary of $A$ to obtain a disk $D'$ with boundary $C_1$, where $D' = D \cup A$. Identify the edges of $D'$, in pairs, to obtain a surface $M(n)$ that contains $D$. The special case where the identification of the edges of $D'$ produces a torus is illustrated in Fig. 5.

The third step in the proof is to notice that Theorem 1 can be re-stated as follows:

ALTERNATIVE STATEMENT OF THEOREM 1. *If $M$ is a surface, different from a sphere, then $M = M_1 \# M_2 \# \cdots \# M_j$, where for each $i$, $M_i$ is either a projective plane or a torus.*

The fourth, and final, step is a proof of this alternative statement, using induction on the number of edges in the disk $D$ obtained in the first step. Let $n$ be a positive even integer such that for any even integer $k$, where $2 \leqslant k < n$, any surface $M(k)$ resulting from identifying, in pairs, the edges of a disk $D(k)$ is a connected sum as required in the conclusion of the alternative statement of Theorem 1. (See the description above of the notation $D(k)$ and $M(k)$.) We wish to show that any surface $M(n)$ is such a connected sum. Let $D(n)$ be a disk with the edges identified to obtain $M(n)$. We assume, by Lemma 1, that $n \geqslant 4$. The inductive argument is separated into four cases, with some overlapping among them.

CASE 1. There is a twisted pair in the identification of the edges of $D(n)$. Identify the two edges in such a twisted pair to obtain a Möbius band $B$ with $n - 2$ edges on its boundary. There is an annulus $A$ in $B$ such that one component of the boundary of $A$ is the boundary of $B$. (See Fig. 6 for an illustration of this where $n = 6$.) Let $J$ denote the other component of the boundary of $A$. Notice that the closure of $B - A$ is a Möbius band and that $J$ is its boundary. Identify the edges in the boundary of $B$ as specified for $D(n)$. By Lemma 2, it is now easy to see that $M(n) = P \# M(n-2)$, where $J$ becomes the identified boundaries of the two disks that are removed to obtain the connected sum of $P$ and $M(n-2)$.

CASE 2. There are two separated pairs of edges of $D(n)$ that are nontwisted. If $n = 4$, then $M(n)$ is a torus. (See Fig. 3.) If $n \geqslant 6$, it is easy to see, by use of Fig. 7, that a punctured torus results from identifying the edges in two separated pairs of edges that are nontwisted. By Lemma 2, $M(n) = T \# M(n-4)$.

CASE 3. There is a nontwisted pair of adjacent edges in $D(n)$. A disk $D(n-2)$ is obtained by identifying these two adjacent edges as indicated in Fig. 8. Thus $M(n) = M(n-2)$.

CASE 4. There is a nontwisted pair of nonadjacent edges of $D(n)$ that does not separate any other identified pair. An annulus $A$ is obtained by identifying the edges in some such nontwisted pair, as indicated in Fig. 9. It is easy to see that $n \geqslant 6$. There are two positive even integers $h$ and $l$ such that $h + l = n - 2$, where $h$ denotes the number of edges in one component of the boundary of $A$ and $l$ the number in the other component. Each edge in each of these two

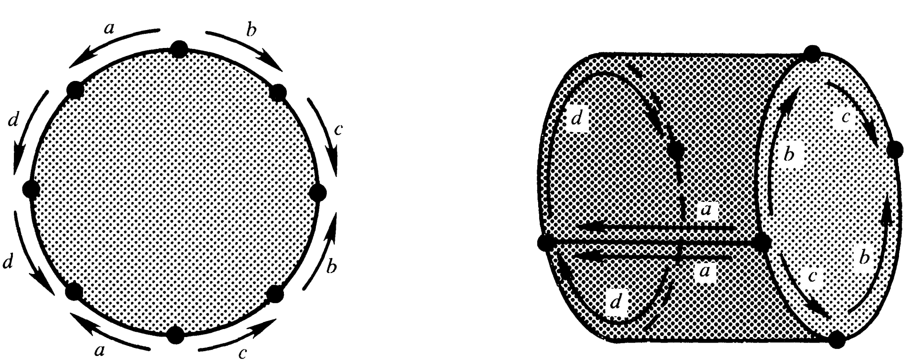


FIG. 9

components must be identified with an edge in the same component. By Lemma 2,

$$M(n) = M(h) \# M(l) = M(n-2).$$

REMARKS. The one-sided property of a Möbius band can be used to show that the connected sum of three projective planes is topologically the same as the connected sum of a torus and a projective plane. This can be used, as in [3, pp. 26–29], to obtain the following more specific classification of surfaces.

THEOREM 2. *Any surface different from a sphere is either a connected sum of a finite number of tori or a connected sum of a finite number of projective planes.*

The inductive procedure in the proof of Theorem 1, combined with Theorem 2, furnishes the following information about the number and types of surfaces that can be obtained from the various identifications of the edges of a given disk.

If $k$ is odd, then $(3k+1)/2$ topologically distinct surfaces can be obtained from $D(2k)$. The connected sum of $k$ projective planes is the only such surface that cannot be obtained from $D(2k-2)$, where $k > 1$.

If $k$ is even, then $(3k+2)/2$ topologically distinct surfaces can be obtained from $D(2k)$. In this case, both the connected sum of $k$ projective planes and the connected sum of $k/2$ tori can be obtained from $D(2k)$ but not from $D(2k-2)$.

### References


1. Michael Henle, A Combinatorial Introduction to Topology,,Freeman, San Francisco, 1979, Chapt. 4.
2. Donald W. Kahn, Topology: An Introduction to Point-set and Algebraic Areas, Williams and Wilkins, Baltimore, 1975, Chapt. 7.
3. William S. Massey, Algebraic Topology: An Introduction, Harcourt, Brace and World, New York, 1967, Chapt. 1.
4. H. Seifert and W. Threlfall, A Textbook of Topology (translated by Michael A. Goldman), Academic Press, New York, 1980, Chapt. 6.


## AN INEQUALITY CONCERNING MINORS OF A SEMIDEFINITE MATRIX

BORIS KIMELFELD
*Konakovsky pr. 15, apt. 23, Moscow 125565, USSR*

Let $S = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$ be a Hermitian $n$ by $n$ matrix of complex numbers, where $A$, $B$, and $C$ are its submatrices of size $m$ by $m$, $m$ by $(n-m)$, and $(n-m)$ by $(n-m)$, respectively $(0 < m < n)$. Here $M^*$ denotes the complex conjugate of the transpose of a matrix $M$.

THEOREM. *If $S$ is semidefinite (either positive or negative) then*

$$|\det S| \leqslant |\det A \cdot \det C|. \tag{1}$$

*For definite $S$ equality in* (1) *holds if and only if $B = 0$.*

Inequality (1) was proved originally by E. F. Beckenbach and R. Bellman in [1], Chapter 2, § 10 and 14, but their proof is based on a representation of the determinant as a multiple integral. In this note we give an alternative proof of (1) which is purely algebraic.

REMARK. The example of $S = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ with $m = 2$ shows that the word "definite" in the last assertion of the Theorem cannot be replaced by "semidefinite".

Using the Theorem for $m = 1$, and proceeding inductively, one derives the following well-known corollary.

COROLLARY. *If $S$ is semidefinite, then $|\det S|$ is less than or equal to the absolute value of the product of all elements of the main diagonal of $S$. For definite $S$ the equality occurs if and only if $S$ is diagonal.*

*Proof of the theorem.* Since (1) holds trivially when $\det S = 0$, and both sides of (1) do not change when $S$ is replaced by $-S$, we may restrict ourselves to the case when $S$ is positive definite. In this case $C$ and, consequently, $C^{-1}$ are positive definite, and $S = U^*\hat{S}U$, where

$$U = \begin{pmatrix} E_m & 0 \\ C^{-1}B^* & E_{n-m} \end{pmatrix}, \qquad \hat{S} = \begin{pmatrix} \hat{A} & 0 \\ 0 & C \end{pmatrix}, \quad \text{and} \quad \hat{A} = A - BC^{-1}B^*.$$

(Here $E_k$ denotes the identity matrix of order $k$.) Since $\det U = 1$, we have $\det S = \det \hat{S} = \det \hat{A} \cdot \det C$. So (1) is equivalent to the inequality $\det \hat{A} \leqslant \det A$, and equality in (1) holds if and only if $\det \hat{A} = \det A$. Positive definiteness of $S$ implies that both $A$ and $\hat{A}$ are positive definite and, consequently, can be reduced simultaneously to the diagonal form, meaning that there exists an $m$ by $m$ matrix $Q$ with $\det Q = 1$ such that $Q^*AQ = \operatorname{diag}(a_1, a_2, \ldots, a_m)$ and $Q^*\hat{A}Q = \operatorname{diag}(\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_m)$. (In [**2**], Chapter 4, Sect. 12, this fact is proved for real symmetric matrices, but the proof holds equally for Hermitian matrices.) The differences $a_i - \hat{a}_i$ $(i = 1, 2, \ldots, m)$ are the eigenvalues of the Hermitian matrix $R = Q^*(BC^{-1}B^*)Q = (B^*Q)^*C^{-1}(B^*Q)$, which is positive semidefinite because of the positive definiteness of $C^{-1}$. Hence $a_i \geqslant \hat{a}_i > 0$ for all $i = 1, 2, \ldots, m$, and therefore

$$\det A = \det Q^*AQ = a_1 a_2 \cdots a_m \geqslant \hat{a}_1 \hat{a}_2 \cdots \hat{a}_m = \det Q^*\hat{A}Q = \det \hat{A}.$$

This proves (1).

Now we see that the sign of equality in (1) holds if and only if all eigenvalues $(a_i - \hat{a}_i)$ of the Hermitian matrix $R$ vanish, that is, if and only if $R = 0$. It remains to show that $R = (B^*Q)^*C^{-1}(B^*Q) = 0$ only if $B = 0$. To prove this, suppose that $B \neq 0$, so $B^* \neq 0$, and there exists an $(n-m)$-dimensional column-vector $x$ such that $y = B^*x \neq 0$. Letting $z = Q^{-1}x$, we have $z^*Rz = y^*C^{-1}y > 0$ because of the positive definiteness of $C^{-1}$, and consequently $R \neq 0$. This completes the proof.

### References

**1.** E. F. Beckenbach and R. Bellman, Inequalities, Springer-Verlag, Berlin, 1961.

**2.** R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, New York, 1970.

---

## MISCELLANEA

**147.** **The Mathematics of Fiction**

[The writer] must think out completely, as coolly as any critic, what his fiction means, or is trying to mean. He must complete his equations, think out the subtlest implications of what he's said, get at the truth not just of his characters and action but also of his fiction's form, remembering that neatness can be carried too far, so that the work begins to seem fussy and overwrought, anal compulsive, unspontaneous, and remembering that, on the other hand, mess is no adequate alternative. He must think as cleanly as a mathematician, but he must also know by intuition when to sacrifice precision for some higher good, how to simplify, take short cuts, keep the foreground up there in front and the background back.

-John Gardner, *The Art of Fiction*,
Alfred A. Knopf, New York, 1983, p. 7.

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## A HEURISTIC FOR THE POISSON INTEGRAL FOR THE HALF PLANE AND SOME CAVEATS

RIDGLEY LANGE
*Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859*

ROBERT A. WALSH
*1203 1/2 Washington Avenue, Golden, CO 80401*

Let $H(x, y)$ be a real-valued function depending on the real variables $x$ and $y$ defined on the upper half plane $y > 0$. Let $\Delta H$ denote the Laplacian $H = H_{xx} + H_{yy}$. In this note we want to indicate an intuitive solution to the Dirichlet problem on the half plane accessible to a student who has had a one semester course in complex variable. Specifically, we give a new heuristic to derive the Poisson integral solution of the boundary value problem

$$\Delta H(x, y) = 0, \qquad -\infty < x < \infty;\ y > 0, \tag{1a}$$

$$H(x, 0) = f(x), \qquad -\infty < x < \infty, \tag{1b}$$

where $f(x)$ denotes the boundary values of the (unknown) harmonic function $H$ on $y > 0$. If $f$ is bounded and piecewise continuous on $(-\infty, \infty)$, the well-known solution $\hat{H}(x, y)$ to (1) at $(x, y)$ is given by the integral formula ($y > 0$)

$$\hat{H}(x, y) = \frac{y}{\pi}\int_{-\infty}^{\infty} \frac{f(\xi)\, d\xi}{(x-\xi)^2 + y^2}. \tag{2}$$

We show how to construct (2) from simple cases of Problem (1) whose solutions can be verified by inspection.

For example, suppose boundary values (1b) are defined by the step function $f = 1$ on $(-\infty, 0]$ and $f = 0$ on $(0, \infty)$. For $(x, y)$ in the upper half plane, it is easy to see that $\hat{H}(x, y) = \frac{1}{\pi}\arg(x + iy)$ is harmonic and satisfies the boundary condition. If $a$ is any real number and $f = 1$ on $(-\infty, a]$ and $f = 0$ on $(a, \infty)$, then the corresponding solution is

$$\hat{H}(x, y) = \frac{1}{\pi}\arg(z - a), \tag{3}$$

where we write $z = x + iy$. Next suppose $f = 1$ on $[a, b]$ and vanishes off $[a, b]$. The linearity of $\Delta$ implies that the solution for (1) is

$$\hat{H}(x, y) = \frac{1}{\pi}[\arg(z - b) - \arg(z - a)]. \tag{4}$$

Equation (4) is our basic building block for the heuristic for Equation (2).

Now let $f$ be bounded and piecewise continuous on $(-\infty, \infty)$ and fix $z = x + iy$, $y > 0$. Let $A < 0 < B$ and partition $[A, B]$ into a subintervals $[\xi_{j-1}, \xi_j]$, $j = 1, 2, \ldots, n$, $\xi_0 < \xi_1 < \cdots < \xi_n$, such that no such interval contains a discontinuity of $f$ in its interior. Let $f(\hat{\xi}_j)$ be a constant approximation to $f$ on the $j$th interval. In this way we approximate $f$ on $[A, B]$ by a sum of $n$ step functions. Using the linearity of $\Delta$ again, we may reasonably say that the Riemann-Stieltjes sum

$$\frac{1}{\pi}\sum_{j=1}^{n} f(\hat{\xi}_j)\left[\arg(z - \xi_{j-1}) - \arg(z - \xi_j)\right] \tag{5}$$

approximates the desired solution to (1), where at this stage we simply ignore the boundary values off $[A, B]$ for $A$ and $B$ sufficiently far from zero. Letting $n \to \infty$ in the sum (5) gives the integral approximation

$$\hat{H}(x, y) \approx \frac{1}{\pi} \int_A^B f(\xi)\, d(\arg(z - \xi)). \tag{6}$$

Assuming further that our error is due to the finiteness of $A$ and $B$, we predict the exact solution to be

$$\hat{H}(x, y) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(\xi)\, d(\arg(z - \xi)). \tag{7}$$

To see that this leads to (2), we perform the following computation using the usual principal branch of the logarithm:

$$\begin{aligned} d(\arg(z - \xi)) &= \frac{d}{d\xi} \arg(z - \xi)\, d\xi \\ &= \operatorname{Im}\left( \frac{d}{d\xi} \ln(z - \xi) \right) d\xi \\ &= \operatorname{Im}\left( \frac{1}{\xi - z} \right) d\xi. \end{aligned} \tag{8}$$

Now (8) can be rewritten as

$$d(\arg(z - \xi)) = \frac{y\, d\xi}{(x - \xi)^2 + y^2}, \tag{9}$$

and (2) follows from (7) and (9).

As a function of $z$ the integral over $[A, B]$ in (6) is, according to (8), the imaginary part of the analytic function

$$g(z) = \frac{1}{\pi} \int_A^B \frac{f(\xi)\, d\xi}{\xi - z},$$

hence (6) represents a harmonic function on $y > 0$. Moreover, an easy estimate by means of the boundedness of $f$ on $(-\infty, \infty)$ and identity (9) shows that the integrals (6) actually converge to integral (2) *uniformly* on discs in $z$ in $y > 0$. This suggests that (2) represents a *harmonic* solution to Problem (1).

Tacit in this last heuristic step is that a system of harmonic functions converging uniformly on neighborhoods of their common domain have a harmonic limit. This rather vague description can be made precise enough to give a proof that (2) is a harmonic solution. From the more familiar Poisson integral for a *disc*, it follows that $h$ is harmonic on an open region $D$ if and only if it satisfies the mean value property, i.e.,

$$h(z_0) = \frac{1}{2\pi} \int_0^{2\pi} h\left(z_0 + re^{i\theta}\right) d\theta \tag{10}$$

on each disc $|z - z_0| \leqslant r$ contained in $D$. (See [**1**, p. 234], for details.) It is evident that the mean value property is preserved under uniform convergence, thus the existence of solution (2) as a harmonic one can be established.

To complete a rigorous solution, it must also be proved that (2) represents the only *bounded* solution to (1). (There may be infinitely many unbounded solutions, e.g., if $f \equiv 0$ in (1b), then $H_1(x, y) = y$ and $H_2(x, y) = 2y$ are distinct unbounded solutions to (1).) The uniqueness of the bounded solution may be established along the lines of [**1**], Exercise 2, p. 169, with the help of (10), and the elementary facts that a nonconstant harmonic function is nonconstant on each disc and has no local extrema on its domain.

We close with some cautionary remarks about boundary behavior on the half plane. Using

representation (2) with elementary proofs, one can show that at each point $x_0$ of continuity of $f$ in (1b)

$$\lim_{y \to 0^+} H(x_0, y) = f(x_0)$$

and for a jump discontinuity of $f$ at $x_0$,

$$\lim_{y \to 0^+} H(x_0, y) = \frac{1}{2}[f(x_{0^+}) + f(x_{0^-})] \tag{11}$$

so the boundary limit is the average of the left and right-hand limits of $f$ at $x_0$. Property (11) looks like one students learn in elementary Fourier expansions, but some new subleties are worth pointing out. For example, the limit in Equation (11) must be *radial*, or normal to the boundary; consider Equation (3) above with $a = 0$. Here (11) becomes (as we expect)

$$\lim_{y \to 0^+} \hat{H}(0, y) = \lim_{y \to 0^+} \frac{1}{\pi} \arg(iy) = \frac{1}{2};$$

but if $(x, y) \to (0, 0)$ along the diagonal $x = y$, we get

$$\lim_{(x, y) \to (0,0)} \hat{H}(x, y) = \frac{1}{\pi} \lim_{x \to 0^+} \arg(x + xi) = \frac{1}{4}.$$

For a discontinuity at $x_0$ which is not a jump, the limit $\lim_{y \to 0^+} H(x_0, y)$ need not have a "reasonable" value. A student might be tempted to guess the radial limit at $x_0 = 0$ for the boundary function $f(\xi) = \sin^2(1/\xi)$ for $\xi > 0$ and $f(\xi) = 0$ for $\xi \leqslant 0$. By (2) if $y > 0$,

$$\hat{H}(0, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{\sin^2(1/\xi)}{\xi^2 + y^2} d\xi,$$

and the last integral can be written as

$$\hat{H}(0, y) = \frac{2}{\pi} \int_0^{\infty} \frac{dt}{t^2 + 1} - \frac{2}{\pi} \int_0^{\infty} \frac{\cos(2t/y)\, dt}{t^2 + 1}. \tag{12}$$

The value of the first integral in (12) is clearly 1, while the second can be found by contour integration or otherwise to be $e^{-2/y}$. Hence

$$\lim_{y \to 0^+} H(0, y) = 1,$$

a result that "intuition" cannot account for. Moreover, if we take $f(\xi) = \sin^2(1/\xi)$ for $\xi \neq 0$, then the radial limit as $(0, y) \to (0, 0^+)$ is found by a calculation similar to (12) to be $\frac{1}{2}$!

**Reference**

1. L. Ahlfors, Complex Analysis, 2nd ed., McGraw-Hill, New York, 1963.

---

## 148. MISCELLANEA

**Definition**

In fact, I am rather inclined to agree with those who maintain that the moments of greatest creative advancement in science frequently coincide with the introduction of new notions by means of definition.

–Alfred Tarski, "The semantic conception of truth and the foundations of semantics," from *Philosophy and Phenomenological Research*, vol. 4, 1944.

## PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

### ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3088. *Proposed by Solomon W. Golomb, University of Southern California.*

Show that, for every positive integer $n$,

$$\sum_{k=1}^{n} \frac{k \cdot k!}{n^k} \binom{n}{k} = n.$$

E 3089. *Proposed by Eliot T. Jacobson, Ohio University.*

Determine all positive even integers $n$ having the following property: given any integer $b$ with $1 < b < n$ and $\gcd(b, n) = 1$, there is a solution to the congruence $(b-1)x \equiv \frac{n}{2} \pmod{n}$.

E 3090. *Proposed by Alfonso Villani, Seminario Matematico, Catania, Italy.*

Let $f\colon [0, +\infty) \to [0, +\infty)$ be a continuous function. Let $h > 0$ be given and let $g\colon [0, +\infty) \to \mathbb{R}$ be defined by

$$g(x) = \frac{f(x)}{f(x+h) - 1}.$$

(a) Show that if $\liminf_{x \to +\infty} g(x) > 0$, then $\lim_{x \to +\infty} f(x) = 0$.

(b) Show that if $\limsup_{x \to +\infty} g(x) < 0$, then $\lim_{x \to +\infty} f(x) = +\infty$.

(c) Assume that $\lim_{x \to +\infty} g(x) = 0$. Is it possible for $f$ to satisfy both $\liminf_{x \to +\infty} f(x) = 0$ and $\limsup_{x \to +\infty} f(x) = +\infty$?

(d) How, in (a) and (b), can the assumption "$f$ is continuous" be weakened? Can this assumption simply be omitted?

E 3091. *Proposed by Calin P. Popescu, student, Bucharest, Romania.*

Equilateral triangle $ABC$ is inscribed in triangle $XYZ$, with $A$ between $Y$ and $Z$, $B$ between $Z$ and $X$, and $C$ between $X$ and $Y$. Show that $XA + YB + ZC < XY + YZ + ZX$. Is equality possible?

E 3092. *Proposed by Rick Luttmann, Sonoma State University.*

Problem A-6 on the 1982 Putnam Competition was:

"Let $\sigma$ be a bijection of the positive integers, that is, a one-to-one function from $\{1, 2, 3, \dots\}$ onto itself. Let $x_1, x_2, x_3, \dots$ be a sequence of real numbers with the following three properties:

(i) $|x_n|$ is a strictly decreasing function of $n$;
(ii) $|\sigma(n) - n| \cdot |x_n| \to 0$ as $n \to \infty$;
(iii) $\lim_{n\to\infty}\sum_{k=1}^{n} x_k = 1$.

Prove or disprove that these conditions imply that

$$\lim_{n\to\infty} \sum_{k=1}^{n} x_{\sigma(k)} = 1."$$

Solve the problem with (ii) replaced by:

(ii)′ $|\sigma(n) - n| \cdot |x_{\sigma(n)}| \to 0$ as $n \to \infty$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### The Extended Erdős-Mordell Inequality

E 2462 [1974, 281]. *Proposed by Huseyin Demir, Middle East Technical University, Ankara, Turkey.*

Let $P$ be a point interior to the triangle $A_1A_2A_3$. Denote by $R_i$ the distance from $P$ to the vertex $A_i$, and denote by $r_i$ the distance from $P$ to the side $a_i$ opposite to $A_i$. The Erdős-Mordell inequality asserts that

$$R_1 + R_2 + R_3 \geqslant 2(r_1 + r_2 + r_3).$$

Prove that the above inequality holds for every point $P$ in the plane of $A_1A_2A_3$ when we make the interpretation $R_i \geqslant 0$ always and $r_i$ is positive or negative depending on whether $P$ and $A_i$ are on the same side of $a_i$ or on opposite sides.

*Editorial note*: Professor Clayton W. Dodge, Department of Mathematics, University of Maine refereed the "solutions" submitted to this problem in 1974 and found that there were no solutions. Since that time Professor Dodge himself has solved the problem. His solution appears in *Crux Mathematicorum*, vol. 10, no. 9, November 1984, pages 274–281.

### Another Manifestation of the Cauchy-Schwartz Inequality

E 2959 [1982, 498]. *Proposed by Jack Garfunkel, Flushing, New York.*

Triangle $ABC$ is inscribed in a circle. The medians of the triangle intersect at $G$ and are extended to the circle to points $D$, $E$ and $F$. Prove: $AG + BG + CG \leqslant GD + GE + GF$.

*Solution by P. Boente, Universität Mannheim, West Germany.* We prove a generalized form of the proposition, namely: Let $P_1, P_2, \dots, P_m$ be points of an $n$-sphere $S$, and $G$ their center of gravity. If the extensions of $P_iG$ intersect $S$ at $Q_i$, then

$$\sum_{i=1}^{m} P_iG \leqslant \sum_{i=1}^{m} Q_iG.$$

*Proof.* We denote the center of $S$ by $M$, its radius by $r$ and by $(.,.)$ and $\|.\|$ the natural inner product in $\mathbb{R}^n$ and its associated norm, respectively. Then we have:

$$\begin{aligned} mr^2 &= \sum_{i=1}^{m} (P_iM)^2 = \sum_{i=1}^{m} \|\overrightarrow{P_iM}\|^2 \\ &= \sum_{i=1}^{m} \|\overrightarrow{P_iG} + \overrightarrow{GM}\|^2 \\ &= \sum_{i=1}^{m} (P_iG)^2 + 2\left(\sum_{i=1}^{m} \overrightarrow{P_iG}, \overrightarrow{GM}\right) + m(GM)^2 \\ &= m(GM)^2 + \sum_{i=1}^{m} (P_iG)^2, \end{aligned}$$

since $\sum_{i=1}^{m}\overrightarrow{P_iG} = 0$, by the definition of $G$, whence

$$\frac{1}{m}\sum_{i=1}^{m} (P_iG)^2 = r^2 - (GM)^2 = P_jG \cdot Q_jG \quad \text{for all } j,$$

by the chords theorem. Division by $P_jG$ and summation yields

$$\begin{aligned} \sum_{j=1}^{m} GQ_j &= \sum_{j=1}^{m} \frac{1}{m}(P_jG)^{-1}\sum_{i=1}^{m} (P_iG)^2 \\ &= \frac{1}{m}\sum_{i=1}^{m} (P_iG)^{-1}\sum_{i=1}^{m} (P_iG)^2 \\ &\geqslant m\left(\sum_{i=1}^{m} P_iG\right)^{-1}\sum_{i=1}^{m} (P_iG)^2 \\ &\geqslant m\left(\sum_{i=1}^{m} P_iG\right)^{-1}\frac{1}{m}\left(\sum_{i=1}^{m} P_iG\right)^2 \\ &= \sum_{i=1}^{m} P_iG, \end{aligned}$$

where the former inequality is given by the relation between the arithmetic and harmonic mean and the latter one is a well-known application of the inequality of Schwarz.

Also solved by 27 other readers and the proposer. A number of readers referred to E 2505 [1974, 1111; 1976, 59–60]. A number of readers also noted that equality occurs if and only if triangle $ABC$ is equilateral.

### An Inequality for Acute Triangles

E 2968 [1982, 697]. *Proposed by George Tsintsifas, Thessaloniki, Greece.*

The points $A_1', A_2', A_3'$ lie on the sides $A_2A_3, A_3A_1, A_1A_2$ of an acute angle triangle $A_1A_2A_3$, respectively. Show that

$$2\sum a_i' \cos A_i \geqslant \sum a_i \cos A_i,$$

where $a_1, a_2, a_3$ are the sides of the triangle $A_1A_2A_3$ and $a_1', a_2', a_3'$ are the sides of the triangle $A_1'A_2'A_3'$.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* We put

$$A_1'A_2 = p_1,\ A_1'A_3 = q_1,\ A_2'A_3 = p_2,\ A_2'A_1 = q_2,\ A_3'A_1 = p_3,\ A_3'A_2 = q_3.$$

Since the distance between $A_i'$ and $A_j'$ is at least equal to the distance of their orthogonal projections onto $A_iA_j$ we have

$$a_1' \geqslant a_1 - q_3\cos A_2 - p_2\cos A_3,$$

$$a_2' \geqslant a_2 - q_1\cos A_3 - p_3\cos A_1,$$

$$a_3' \geqslant a_3 - q_2\cos A_1 - p_1\cos A_2.$$

Hence, since $\cos A_i > 0$,

$$\Sigma a_i'\cos A_i \geqslant \Sigma a_i\cos A_i - a_1\cos A_2\cos A_3 - a_2\cos A_3\cos A_1 - a_3\cos A_1\cos A_2.$$

With the circumradius of $\Delta A_1A_2A_3$ as the unit of length we find

$$\Sigma a_i\cos A_i = \Sigma\sin 2A_i = 4\sin A_1\sin A_2\sin A_3,$$

and

$$\begin{aligned} a_1\cos A_2\cos A_3 + a_2\cos A_3\cos A_1 + a_3\cos A_1\cos A_2 \\ &= 2\cos A_3(\sin A_1\cos A_2 + \cos A_1\sin A_2) + 2\sin A_3\cos A_1\cos A_2 \\ &= 2\sin A_3(\cos A_3 + \cos A_1\cos A_2) \\ &= 2\sin A_1\sin A_2\sin A_3 = \tfrac{1}{2}\Sigma a_i\cos A_i. \end{aligned}$$

This leads to

$$2\Sigma a_i'\cos A_i \geqslant \Sigma a_i\cos A_i,$$

with equality if and only if in each of the inequalities for $a_1', a_2', a_3'$ equality occurs, that is if and only if $A_1'$, $A_2'$ and $A_3'$ are the midpoints of the sides $A_2A_3, A_3A_1, A_1A_2$, respectively.

Also solved by J. Dou (Spain), I. E. Leonard, W. A. Newcomb, K. L. Stellmacher and the proposer.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by September 30, 1985. The solver's full post-office address should be on each sheet.*

6496. *Proposed by Ryszard Szwarc, University of Wrocław, Poland.*

Let $T$ be a bounded operator with spectral radius 1 on a Hilbert space $H$, and let $c > 1$. Prove that there is an invertible operator $A$ on $H$ such that $\|ATA^{-1}\| \leqslant c$.

6497. *Proposed by Richard Askey, University of Wisconsin.*

Let $0 < q < 1$, $\operatorname{Re} a > 0$ and $\operatorname{Re} b > 0$. Show that

$$\int_0^\infty \frac{(-tq^b;q)_\infty(-q^{a+1}/t;q)_\infty}{(-t;q)_\infty(-q/t;q)_\infty}\frac{d_qt}{t} = \frac{\Gamma_q(a)\Gamma_q(b)}{\Gamma_q(a+b)} \tag{1}$$

and

$$\int_0^\infty \frac{(-tq^b;q)_\infty(-q^{a+1}/t;q)_\infty}{(-t;q)_\infty(-q/t;q)_\infty}\frac{dt}{t} = \frac{-\log q}{1-q}\frac{\Gamma_q(a)\Gamma_q(b)}{\Gamma_q(a+b)}, \tag{2}$$

where

$$(x;q)_\infty := \prod_{n=0}^{\infty}(1-xq^n),$$

$$\Gamma_q(x) := (q;q)_\infty (1-q)^{1-x}/(q^x;q)_\infty,$$

and

$$\int_0^\infty f(t)\,d_q t := (1-q)\sum_{-\infty}^{\infty} f(q^n)q^n.$$

These extend the gamma function identity

$$\int_0^\infty \frac{dt}{t(1+t)^b(1+t^{-1})^a} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

to $q$-gamma functions (for properties see R. Askey, Ramanujan's extensions of the gamma and beta functions, this MONTHLY, 87 (1980) 346–359).

## SOLUTIONS OF ADVANCED PROBLEMS

### Degrees of Irreducible Polynomials over a Field

6043 [1975, 766; 1977, 304]. *Proposed by Brian Peterson, University of California, Berkeley.*

Let $P$ be a nonempty proper subset of the primes. Consider algebraic extensions $F$ of the rationals $Q$ with the property

(*) Every $x$ in $F$ has degree over $Q$ divisible only by primes in $P$.

A Zorn's lemma argument shows that there exist maximal extensions satisfying (*). Is such a maximal extension unique up to isomorphism?

*Solution by M. J. Pelling, University College, London.* The published solution (1977, 304) is invalid since if $u = 10u_1 = 10(2^3 5^3 + 5^3 v)^{1/2}$, then $v = -2^3 + u^2/10^2 5^3 \in Q(u)$ contrary to assertion. One can even show that the solver's method never works for $P = \{5\}$ since it is impossible to have two irreducible quintics $f(x), g(x)$ over $Q$ such that $g(x)$ factors over $Q(u)$, where $u$ is a root of $f(x)$, into the product of an irreducible quadratic and an irreducible cubic.

However, the idea can be made to work for $P = \{7\}$, since there do exist irreducible polynomials $f(x), g(x)$ over $Q$ of degree 7 such that $g = g_3 g_4$ factors into the product of an irreducible cubic $g_3(x)$, and an irreducible quartic $g_4(x)$ over $Q(u)$ where $u$ is a root of $f(x)$.

To prove this let $G$ be the simple group of order 168 which may be obtained as the Galois group of a suitable irreducible polynomial $f(x)$ over $Q$ of degree 7, with $G$ acting as a doubly transitive permutation group on the roots $u_1, \dots, u_7$ of $f(x)$. It is known that $G$ can be obtained as a doubly transitive permutation group on 7 symbols in essentially one way only, so that with the roots of $f(x)$ suitably numbered the 7 quantities

$$\begin{aligned}
v_1 &= u_5 + u_6 + u_7 & v_4 &= u_2 + u_3 + u_5\\
v_2 &= u_3 + u_4 + u_7 & v_5 &= u_1 + u_4 + u_5\\
v_3 &= u_2 + u_4 + u_6 & v_6 &= u_1 + u_3 + u_6\\
& & v_7 &= u_1 + u_2 + u_7
\end{aligned}$$

are permuted (doubly transitively) amongst themselves by the action of $G$. The $v_i$ are therefore the roots of an irreducible polynomial $g(x)$ over $Q$ of degree 7, and since the subgroup of $G$ stabilising $u_1$ has, acting on the $v_i$, orbits $\{v_5, v_6, v_7\}$ and $\{v_1, v_2, v_3, v_4\}$, it follows that $v_5, v_6, v_7$ are the roots of an irreducible cubic $g_3(x)$ and $v_1, v_2, v_3, v_4$ are the roots of an irreducible quartic

$g_4(x)$ over $Q(u_1)$. Hence a maximal extension of $Q$ satisfying (*) for $P = \{7\}$ and containing a root of $f(x)$ cannot be isomorphic to a similar extension containing a root of $g(x)$.

There are also choices of $P$ for which the maximal extension $F$ is unique up to isomorphism: obviously this happens if $P$ is the set of all primes, since then $F$ is the algebraic closure of $Q$, but it also happens if $P$ omits just one prime $p$. To prove this, suppose $F, F'$ are two maximal extensions with isomorphic subfields $K, K'$, respectively, and let $t \in F \setminus K$ so that $t$ will satisfy an irreducible polynomial $h(x)$ over $K$ of degree not divisible by $p$. If $s: K \to K'$ is the isomorphism, the conjugate polynomial $h^s(x)$ is irreducible over $K'$ and must have a root $t'$ in $F'$: for if $h^s = h_1 h_2 \cdots h_k$ is the factorization of $h^s$ into irreducibles over $F'$, then some $h_t$ has degree not divisible by $p$, but if $\deg(h_t) > 1$, we could adjoin a root of $h_t$ to $F'$ and get a larger field satisfying (*) so contradicting maximality of $F'$.

The isomorphism $s$ can therefore be extended to one of $K(t)$ and $K(t')$, mapping $t$ to $t'$, and from this an application of Zorn's lemma shows that $F, F'$ are themselves isomorphic.

### The Derivatives of $x^x$

6392 [1982, 429; 1984, 60; 1984, 372]. *Proposed by C. Ward Henson, Bruce Reznick and Lee A. Rubel, University of Illinois.*

Is there some $x_0 > 0$ such that if $f(x) = x^x$, then $f^{(n)}(x) \geqslant 0$ for all $x \geqslant x_0$ and all $n = 0, 1, 2, \ldots$?

**Apology.** The original published solution by Paul R. Chernoff [1984, 60] was mistakenly judged to be incorrect [1984, 372]. The editors wish to apologize to Paul Chernoff and to thank Bertram Walsh for pointing out that Chernoff's solution is valid since the Taylor series of $f(x) = x^x$ about $x = x_0 > 0$ in fact converges to $f(x)$ in a neighbourhood of $x_0$.

### Stability of Subspaces under Normal Linear Transformations

6432 [1983, 402]. *Proposed by Michael Barr, McGill University.*

Problem 4 on p. 348 of Jacobson's *Basic Algebra I* reads, "Call a linear transformation *normal* if it commutes with its adjoint [with respect to a symmetric or alternating, nonsingular bilinear form on the vector space $V$—*not* its classical adjoint]. ... Show that if $U$ [a subspace of $V$] is stabilized by a normal linear transformation $T$, then $U^\perp$ is stabilized by $T$."

(a) Show by example that the assertion is false.

(b) Show that if the form is anisotropic on a finite dimensional space (no nonzero vector is orthogonal to itself—this requires the form to be symmetric) and the ground field is perfect, then $T$ is diagonalizable over the algebraic closure of the ground field.

(c)* Is it the case that if $U$ is stabilized by a linear transformation $T$ normal with respect to an anisotropic form (over a perfect field, if necessary), then $U^\perp$ is stabilized by $T$?

*Solution to part* (a) *by the proposer.* Let the bilinear form be given by $\langle u, v\rangle = u^t A v$, where $A$ is a nonsingular $n \times n$ matrix and $u$ and $v$ are $n$-dimensional column vectors. Take $n = 4$ and

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \alpha & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

where $\alpha$ is chosen as 1 or $-1$ according as the example is to be symmetric or alternating. Since $A^{-1} = \alpha A$, the adjoint $T^* = \alpha A T^t A = \alpha ATA$ so that normality reduces to $ATAT = TATA$. Since $T^{-1} = T$ and $A^{-1} = \alpha A$, this reduces to $(AT)^4 = I$ which is easily calculated. Thus $T$ is normal. If $\{v_1, v_2, v_3, v_4\}$ is the standard basis, then $(v_1)^\perp = (v_1, v_2, v_4)$ and $T(v_1) \subseteq (v_1)$ while $T(v_1)^\perp = (v_1, v_2, v_3) \not\subseteq (v_1)^\perp$.

*Solution to parts* (b) *and* (c)* *by Daniel Zelinsky, Northwestern University, Evanston, Illinois.*

(b) According to the usual theory, if the minimum polynomial of $T$ is $\Pi p_i^{e_i}$ where the $p_i$'s are distinct, monic, irreducible polynomials over the ground field $F$, then $V$ is the direct sum of $T$-stable subspaces

$$W_i = \{v \in V | p_i(T)^{e_i}(v) = 0\}$$

and each $W_i$ is necessarily stable under any transformation that commutes with $T$ (e.g., under $T^*$, if $T$ is normal). We prove that if $T$ is normal, then every $e_i$ is 1. If not, pick $f$ such that $e_i/2 \leqslant f < e_i$, and write $q = p_i^f$. Then $p_i^{e_i}$ divides $q^2$, so $q(T)^2$ acts as 0 on $W_i$. But the minimum polynomial of the restriction of $T$ to $W_i$ is $p_i^{e_i}$, which does not divide $q$, so $q(T)$ is not zero on $W_i$. The defining property of $T^*$ implies $(u, q(T^*)v) = (q(T)u, v)$ for all $u, v$ in $V$, so, for all $v$ in $W_i$,

$$(q(T)q(T^*)v, q(T^*)q(T)v) = (q(T)q(T)q(T^*)v, q(T)v) = 0$$

because $q(T^*)v \in W_i$ (since $W_i$ is $T^*$-stable) and $q(T)^2$ is zero on $W_i$. By the anisotropic property and $q(T^*)q(T) = q(T)q(T^*)$ (which follows from $TT^* = T^*T$), we know $q(T^*)q(T)v = 0$. Now do it again:

$$(q(T)v, q(T)v) = (v, q(T^*)q(T)v) = (v, 0) = 0,$$

so $q(T)v = 0$. Hence $q(T)$ is zero on $W_i$, a contradiction. This proves $e_i = 1$. Since the ground field $F$ is perfect, it now follows that the minimum polynomial of $T$ factors as a product of distinct linear factors over the algebraic closure $\bar{F}$ of $F$. Thus $T$ is diagonalizable over $\bar{F}$.

(c)* Let $V$ be the space of 4-tuples with entries in the field $F$ of rational numbers. Define linear transformations on $V$ by

$$T([x, y, z, t]) = [y, -x, t, -z] \quad \text{and} \quad S([x, y, z, t]) = [y, -x, -t, z].$$

Then $ST = TS$, but there are $T$-stable subspaces that are not $S$-stable, for example ($[x, y, x, y]$: $x, y \in F$). If we construct an anisotropic inner product on $V$ for which $S = T^*$, we shall have a counterexample to (a) and (c) because $U^\perp$ is not stabilized by $T$ if $U$ is not stabilized by $T^*(= S)$.

Define the inner product of $[x, y, z, t]$ and $[x', y', z', t']$ to be

$$xx' - yy' + 3xy' + 3yx' + 7zz' + 7tt'.$$

It is easy to check that $S = T^*$, that is, $(T(u), v) = (u, S(v))$ for all 4-tuples $u$ and $v$. This inner product is anisotropic because $x^2 - y^2 + 6xy + 7z^2 + 7t^2$ cannot be zero for rational numbers $x, y, z, t$ that are not all zero. If it were, then by clearing denominators, the same quadratic form would be zero for integers $x, y, z, t$ with no common factor. Then

$$(x + 3y)^2 - 10y^2 \equiv 0 \pmod 7.$$

But $-10$ is not a quadratic residue mod 7, so $x + 3y$ and $y$ are both divisible by 7; hence $x$ and $y$ are. Dividing $x^2 - y^2 + 6xy + 7z^2 + 7t^2$ by 7, then reading again mod 7, we get

$$z^2 + t^2 \equiv 0 \pmod 7)$$

and, since $-1$ is not a quadratic residue either, both $z$ and $t$ are divisible by 7. Thus $x, y, z, t$ all have a common factor 7, which is a contradiction.

Parts (a) and (b) were also solved by Edward T. Wong. Note that the counterexample to part (c)* affords another counterexample to part (a).

### Avoiding Meromorphic Functions

6437 [1983, 485]. *Proposed by Bhaskar Bagchi, Indian Statistical Institute, Calcutta.*

Given $n$ meromorphic functions $f_1, f_2, \ldots, f_n$ on a planar region $U$, show that there exists a

holomorphic function $g$ on $U$ such that $f_j(z) \neq g(z)$ for $1 \leqslant j \leqslant n$ and for all $z$ in $U$.

*Solution for the case $n = 1$ by the proposer.* Given a meromorphic function $f$ on $U$, we can write $f = f_1/f_2$ where $f_1, f_2$ are holomorphic with no common zeros. Let $A$ be the set of all zeros of $f_2$. For $\alpha \in A$, let $n(\alpha)$ be the order of the zero of $f_2$ at $\alpha$. It suffices to show the existence of a *nonvanishing* holomorphic function $h$ on $U$ such that

$$h^{(k)}(\alpha) = f_1^{(k)}(\alpha) \quad \text{for} \quad 0 \leqslant k \leqslant n(\alpha),\ \alpha \in A. \tag{1}$$

For then $g = (f_1 - h)/f_2$ is holomorphic on $U$ and $f - g = h/f_2$ is nonvanishing.

An easy computation shows (since $f_1(\alpha) \neq 0$ for $\alpha \in A$) that there are complex numbers $a_k(\alpha)$ for $0 \leqslant k \leqslant n(\alpha)$, $\alpha \in A$, such that whenever a holomorphic function $h_1$ satisfies

$$h_1^{(k)}(\alpha) = a_k(\alpha) \quad \text{for} \quad 0 \leqslant k \leqslant n(\alpha),\ \alpha \in A, \tag{2}$$

then $h = \exp(h_1)$ satisfies (1). But, by a well-known consequence of the Mittag-Leffler theorem, there is a holomorphic function $h_1$ on $U$ satisfying (2) (see, for example, Walter Rudin, *Real and Complex Analysis*, Second Edition, McGraw-Hill, p. 327). This completes the solution for the case $n = 1$.

*Observations by Lee A. Rubel, University of Illinois, Urbana, Illinois.* The assertion is false in general. In case $U = \mathbb{C}$, let $f_1(z) = 0$, $f_2(z) = 1$, and $f_3(z) = z$. Then if the entire (or even meromorphic) function $g$ has a graph that never hits the graphs of $f_1$ and $f_2$, $g$ must be a constant (by Picard's Little Theorem) and so $g(z_0) = f_3(z_0)$ for suitable $z_0$. In the paper, *Interpolation and unavoidable families of meromorphic functions*, by myself and Chung-Chun Yang, in the Michigan Math. J., 20 (1973) pp. 289–296 (see Theorem 3, p. 294 ff.), we show that no two meromorphic functions $f_1$ and $f_2$ have the desired property. Essentially the same proof works if $U$ is the plane with finitely many punctures.

Also solved by Michael von Renteln (West Germany).

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Lars Valerian Ahlfors: Collected Papers, Volume I, 1929–1955, Volume II, 1954–1979.* By Lars Ahlfors. Contemporary Mathematicians. Edited by Gian-Carlo Rota. Birkhauser, Boston, 1983. Vol. I, xix + 520 pp. $65.00 Vol. II, xix + 515 pp. $65.00

CLIFFORD J. EARLE
*Department of Mathematics, White Hall, Cornell University, Ithaca, NY 14853*

Geometric function theory was born on December 26, 1851, when Riemann presented his inaugural dissertation. Today, more than 130 years later, the subject is still flourishing, thanks in large part to Lars Ahlfors, this century's premier geometric function theorist.

As its name suggests, geometric function theory emphasizes the geometric properties of the mappings defined by complex analytic functions. Thus, a one-to-one analytic function is a conformal map, and arbitrary analytic functions are, at least locally, branched covering maps. From this point of view, for instance, the complex exponential function $w = e^z$ defines a universal covering of the punctured $w$-plane $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ by the complex $z$-plane. The horizontal lines $y = 2n\pi i$, $n$ an integer, divide the $z$-plane into parallel strips, each of which is mapped

conformally onto the complement $D$ of the positive real axis in $\mathbb{C}^*$. The $z$-plane can be reconstructed from infinitely many copies of $D$ by assembling them so that crossing the positive real axis corresponds to moving from one of the parallel strips to one of its neighbors. This point of view has far-reaching consequences, leading inevitably to the introduction of Riemann surfaces and analytic mappings between them.

The best way to convey the flavor of modern geometric function theory is to discuss some of Ahlfors' theorems, and the obvious place to start is with the Ahlfors theory of covering surfaces, for which he received the Fields medal in 1936. Let $f$ be a nonconstant analytic function defined in the entire plane, and let $D$ be any region in the plane. A bounded plane region $G$ such that $f$ maps $G$ onto $D$ in an $n$-to-one manner (counting multiplicities) is called an island over $D$, and $n$ is called its multiplicity. Thus, if $G$ is an island of multiplicity $n$, the map $f\colon G \to D$ is a branched covering of degree $n$, and if $n = 1$, then $f$ is a conformal map of $G$ onto $D$. If $f$ omits some value $w_0$, there can be no islands over any region that contains $w_0$. For instance, if $f$ is the exponential function, there are no islands over the unit disk. A more interesting example is $f(z) = \sin z$. This function has the property that $f'(z) = 0$ if $f(z) = \pm 1$, so there are no islands of multiplicity one over the disks $\{w;\ |w - 1| < \frac{1}{2}\}$ or $\{w;\ |w + 1| < \frac{1}{2}\}$.

We wish to state two theorems of Ahlfors, which show that these examples represent the extreme case. The first theorem states that if the regions $D_1$ and $D_2$ are bounded, simply connected, and have disjoint closures, then there is an island over at least one of them. This result immediately implies Picard's theorem that a nonconstant analytic function defined in the entire plane can omit at most one value. The second theorem is even more remarkable. It states that if we are given three bounded simply connected regions with disjoint closures, there must be an island *of multiplicity one* over at least one of them. The Ahlfors theory deals with more general situations, but these two theorems should be sufficient to illustrate its power.

The use of ideas from differential geometry in geometric function theory is well illustrated by the Ahlfors extension of Schwarz's lemma. The classical Schwarz lemma says that if the analytic function $f(z)$ maps the open unit disk $U$ into itself and $f(0) = 0$, then $|f(z)| \le |z|$ for all $z$ in $U$. In 1916 G. Pick observed that the special role of the point $z = 0$ can be eliminated by using the hyperbolic (non-Euclidean) distance

$$d(z,w) = \tanh^{-1}\left|\frac{z - w}{1 - \bar{z}w}\right|$$

in $U$. Pick's version of the Schwarz lemma states that $d(f(z_1), f(z_2)) \le d(z_1, z_2)$ whenever $f(z)$ is an analytic function that maps $U$ into itself. Since the hyperbolic distance comes from the Riemannian metric $ds = (1 - |z|^2)^{-1}|dz|$ on $U$, Pick's theorem is equivalent to the inequality

$$|f'(z)|\left(1 - |f(z)|^2\right)^{-1} \le \left(1 - |z|^2\right)^{-1},$$

valid whenever $f\colon U \to U$ is analytic. Ahlfors interprets that inequality as a statement about metrics of negative curvature in the following way. If $\lambda(z)$ is a positive smooth function in some region, the (scalar) curvature of the Riemannian metric $ds = \lambda(z)|dz|$ is defined by the formula

$$K(z) = -\lambda(z)^{-2}\Delta(\log\lambda)(z).$$

Here $\Delta$ is the Laplacian. The Ahlfors extension of Schwarz's lemma states that if $\lambda(z)$ is a nonnegative function in $U$ such that $K(z) \le -4$ whenever $\lambda(z) > 0$, then $\lambda(z) \le (1 - |z|^2)^{-1}$ everywhere in $U$. The proof is astonishingly elementary. Pick's inequality follows at once, since

$$\lambda(z) = |f'(z)|\left(1 - |f(z)|^2\right)^{-1}$$

satisfies $K(z) = -4$ when $\lambda(z) > 0$. In other applications the hard step is usually to construct an appropriate metric $\lambda(z)|dz|$. Ahlfors makes this task somewhat easier by reformulating the condition $K(z) \le -4$ in a way that does not require $\lambda(z)$ to be smooth, but such technical refinements need not concern us here.

The fact that these illustrative examples come from the mid-1930's should not lead the reader to conclude that geometric function theory has stood still since then. On the contrary, the 1950's and 1960's saw rapid development in the theory of quasiconformal mappings and their application to the study of Teichmüller spaces. Ahlfors and Bers were the leaders of this activity, and many of the papers collected here deal with these subjects.

In 1964 Ahlfors published a paper about finitely generated Kleinian groups that brought this subject back to life and initiated a growing body of important research. Consider a finitely generated group $G$ of Möbius transformations. Let $D$ be the largest open subset of the Riemann sphere in which $G$ acts properly discontinuously, and assume $D$ is not empty. Then the quotient map $D \to D/G$ is a branched covering, and every connected component of $D/G$ is a Riemann surface. The Ahlfors finiteness theorem asserts that $D/G$ has only finitely many branch points and connected components and that each connected component is either compact or the complement of a finite subset of a compact Riemann surface. With the help of that deep theorem much can be said about the geometry of finitely generated Kleinian groups. More recently, Thurston's theory of 3-manifolds has made Kleinian groups even more important. Much of Ahlfors' recent work has dealt with extending to the $n$-dimensional case the analytic tools that have been so successful in studying these groups in two dimensions. The very last paper in these volumes deals with the proof of an important theorem of Dennis Sullivan about ergodic properties of $n$-dimensional Kleinian groups. It is clear that geometric function theory is still very much alive, and that the papers collected here are central to its modern development.

One unexpected feature of these volumes calls for comment. The editors had the brilliant idea of asking Ahlfors to provide his own running commentary on these collected papers. His comments are always enlightening, and the reader who wants to learn more about geometric function theory would do well to start there. It goes without saying that those who already know and love the subject will find these volumes a treasured and indispensable addition to their libraries.

---

*Political and Related Models*. Edited by Steven J. Brams, William F. Lucas, and Philip D. Straffin, Jr. Modules in Applied Mathematics, Volume 2. Springer-Verlag, New York, 1983. xx + 396 pp.

JOSEPH MALKEVITCH
*Department of Mathematics, York College, Jamaica, NY 11451*

The power of mathematics to solve problems in the physical and biological sciences and in engineering is widely known outside the mathematics community. Unfortunately, however, the mathematical techniques used to solve these problems lie beyond what can be easily described to laymen and nonadvanced students. The standard tools for applications in these areas often involve ordinary and partial differential equations, complex variables, Fourier analysis and special functions. This situation is unfortunate if one believes that it is valuable to teach applications to encourage interest and participation in mathematics by a broader group of people. However, in the years since World War II there has been tremendous growth in using mathematics to solve problems in nontraditional areas such as political science, management science, computer science and psychology. This is noteworthy not merely because persons outside the mathematics community (and to some extent within it) are not aware of these applications, but also because the methods in these applications usually involve discrete mathematics. This makes it possible to teach significant mathematics, motivating the study by novel applications, to students with minimal mathematical preparation. Furthermore, the mathematics these applications lead to is elegant, important, often deep, and gives hope of bringing full circle the delicate interplay of

applications-oriented and pure mathematical developments that are so much a part of the history of mathematics. We hope the books in this series, especially Volume 2, will make more people aware of these exciting developments.

Volume 2 consists of fourteen articles written by people who generally have reputations both as researchers and expositors. The volume focuses primarily on models in political science, with a smattering of other subjects also being treated. The lead article, "The Process of Applied Mathematics" by Maynard Thompson, gives a brief, useful, albeit somewhat "standard" discussion of the modeling process. The article centers on two examples, one from physics and one from psychology, which is unfortunate since the flavor of the applications in political science is rather different.

The strength of this book lies in the material in: Chapter 2 ("Proportional Representation" by Edward Bolger), Chapter 3 ("Comparison Voting" by Steven J. Brams), Chapter 4 ("Modeling Coalitional Values" by William F. Lucas and Louis J. Billera), Chapter 9 ("Measuring Power in Weighted Voting Systems" by William F. Lucas), Chapter 10 ("To the (Minimal Winning) Victors Go the (Equally Divided) Spoils: A New Power Index for Simple *n*-Person Games" by John Deegan, Jr., and Edward W. Packel), Chapter 11 ("Power Indices in Politics" by Philip D. Straffin, Jr.), Chapter 12 ("Committee Decision Making" by Peter Rice), and Chapter 14 ("The Apportionment Problem" by William F. Lucas). These articles, augmented by Chapter 5 ("Urban Wastewater Management Planning" by James P. Heaney), cover a fairly broad range of applications to political science and deal with a rich collection of ideas. The nonpolitical-science articles round out the volume, and though the best of them (for example, Chapter 8, "How to Ask Sensitive Questions Without Getting Punched in the Nose" by John C. Maceli) contribute interesting ideas, the tend to dilute the "unity" of the rest of the book. The "unity problem" was certainly an inevitable consequence of the fact that nearly all the articles which make up the four books in the series arose from two writing projects sponsored by the Mathematical Association of America with support in part from the National Science Foundation. Considering that no attempt was made to coordinate what was originally written, the modules go together remarkably well.

Since the book is aimed at teachers and students, nearly all the chapters have exercises, "notes for the instructor," and bibliographies. The exercises are very well chosen. There are some mechanical problems, but there is particular emphasis on more open ended problems and projects. The union of the references in these papers is very fine and will enable one to explore in many directions. (References are given to both general works and research papers.) The notes to the instructors are of a rather general kind and are unlikely to be of great value. More valuable would have been an index which, unhappily, none of the four volumes has.

Minor quibbles aside, these papers constitute a fresh, well-written, fascinating account of how mathematics is being used to obtain exciting, useful results in political science and other areas. This book deserves to be in the personal library of all mathematicians looking for new teaching materials and intellectual stimulation.

---

*Chess Skill in Man and Machines*. Peter W. Frey (Editor). Second Edition. Texts and Monographs in Computer Science, edited by David Gates. Springer-Verlag, New York, 1983. xiv + 329 pp.

HANS BERLINER

*Computer Science Department, Carnegie-Mellon University, Pittsburgh, PA 15213*

Before saying anything else, it should be noted that this is by far the best book on computer chess. It is the result of a seminar series held at Northwestern University in 1974. This series brought together a number of technical people who were actually developing computer chess programs, and others who were either good chess players or had studied the psychology of chess

playing. Among the former were the outstanding duo at Northwestern, David Slate and Larry Atkin, whose programs dominated computer chess competition during the 1970's. Their account of the workings and development of their program is the cornerstone of this volume. Their chapter is a very detailed exposé of the innards of the best machine at the time, and how it got to be that way. Enough detail is provided to allow a programmer to build a program to mimic large parts of CHESS 4.5 (as the Northwestern program was then called), and this article has, no doubt, been used amply in the construction of the many microprocessor programs that are commercially available now.

There are two other parts of the first edition that I also like very much. One is a chapter by Neil Charness on the psychology of chess and the experimental evidence that supports this. This is also far and away the best article of its kind. The other chapter that I liked was one by the volume editor, Peter Frey, on how to understand the workings of a chess program. This is a patient and thorough treatment that is easy to understand, and one from which the interested layman can learn a lot. This core of three fine papers was supplemented by others that were either not so fine, or which had less lofty aims, or had been published in essentially the same form elsewhere. However, these did add a number of aspects to the whole, including a treatment of the history of computer chess tournaments up to 1975, two fine appendices of games of CHESS 4.5 and 4.6, and an overview of some approaches that did not work out. In any case, the three excellent papers were clearly worth the price of the book.

The second edition contains all of the first plus two new chapters and a large new appendix of computer games. One new chapter is Joe Condon and Ken Thompson's account of how Belle, a once mediocre program, was transformed into the most powerful chess machine of today with the relentless refinement of parts of the program by the addition of special purpose hardware to do the computations faster and in most cases in parallel. This special purpose parallelization has two aims. It allows machines to

(1) Avoid the cost of "general purpose" computation in favor of computations that do the minimal necessary amount of work, and

(2) Where possible, have several machines operating in parallel on different aspects of some task that must be performed, thus avoiding the standard serial costs of computation.

I found this chapter very worthwhile, both technically and as an overview of what must be done to achieve the levels that Belle has reached.

The other new chapter is by David Wilkins and discusses his program that uses plans to guide the search. This is very interesting work which unfortunately cannot be judged at this time, since it was a doctoral dissertation subject which showed promise but has not been pushed further.

Many excellent computer games have now been played. A good sample of these is in the new appendix and offers insights into what the standard type of program, searching much deeper than at the time the volume first came out, can now do and not do. All these games are annotated, some by more than one commentator. For some games, computer printout data is also provided. All this makes them very useful both to interested chess players and computer chess afficionados.

This is an overview of what worthwhile there is to be found. However, it is interesting to examine this book from another perspective: that of the practitioner versus the theorist. From this point of view the theorist (meaning the academic) comes off very badly.

Since people in glass houses shouldn't throw stones, let me start by explaining that I participated in the original symposium, but had already committed my contribution to another volume before the idea for this book was broached. I was clearly among the theorists, though I had built one moderately good chess program (for its time). My paper was one that had a number of "promising" ideas, which were duly exposed, and the virtues of which were extolled. Unfortunately, the program that they were embodied in never played well, though I still cherish some of the ideas and believe they will eventually find their place in the scheme of things.

This general tone, of the academic who must praise his research while either not having the

time or being otherwise unwilling to expose it to the cauldron of test by trial, permeates a number of papers in this volume. This is in contrast with the papers of Slate and Atkin, and Condon and Thompson, both the products of superachievers who consistently downplay their accomplishments. Computer chess is a hard-nosed arena in which to test your skill. Clever ideas are no good unless they can be brought into a well-rounded vehicle that will perform correctly a high percentage of the time. Cleverness fades into mediocrity without this.

Somehow, I see something tragicomic in this. Is it necessary for an academic to overpresent his ideas in order to have people pay attention? Can the academic simply not afford the time to test his product, or is he afraid that such tests will show him to be a rather mediocre pretender to excellence? These appear to be important questions.

Some highly regarded academics whom I know would say that the good work will survive and the chaff will be forgotten. That is certainly a humane and balanced notion. I yearn for the time, however, when it will be possible for an academic to portray the weaknesses of his own work as clearly as the strengths.

---

*Applied Statistics: A Handbook of Techniques*. By Lother Sachs. Springer Series in Statistics. Translated by Zenon Reynarowych. Springer-Verlag, New York, 1982, xvi + 706 pp.

GREG CONSTANTINE
*Department of Mathematics, Indiana University, Bloomington, IN 47405*

You may find this a bit unusual, but the book by L. Sachs contains no proofs; it describes, just as its title suggests, a handful of techniques. It provides us with recipes for performing statistical tests. Should such books be written? Are they of any use to anyone? I can't answer either question with just yes or no.

A book of recipes... That does not entail much intellectual challenge—it's bound to be dull, and dull it is. But then one of the best meals I ever had was prepared by a friend of mine just the other day. Alas, no gourmet cook is my friend. Yet, a veritable gourmet piece was created (believe me, I can tell) by faithfully following a recipe from a cookbook, and on a first try.

Should we really present mathematics in the form of recipes? It depends on the audience. Think of Pólya's Enumeration Theorem. It counts the number of patterns of a certain weight which arise on a set under the action of a group, such as the number of ways one can place 16 checkers on an ordinary chessboard—two arrangements being the *same* if one can be obtained from the other by a rotation of the chessboard. The proof of this theorem entails a knowledge of group theory, generating functions, and a great deal of counting. It is enjoyable for a mathematician to go over it in detail. As is well known, the theorem also has a great number of practical uses. For example, a chemist might on occasion want to use it. Should we then require him to spend lots of time beforehand in order to understand all the details of the proof? Instead, we could just present the theorem to him in the form of a clear recipe. Something (roughly) like this: Ingredients: Set $S$ and group $G$. (1) Find the cycle structure of $G$ as a permutation group on set $S$ (then throw group in garbage can). (2) Obtain from (1) the polynomial $P_G\ (x_1, \dots, x_m)$ called the cycle index. (3) Substitute the number $\Sigma_r\ (w(r))^t$ for $x_t$ in $P_G$—and out pops the number of nonequivalent patterns under $G$ using the weight function $w$ (which is what the chemist wants). It works: don't knock it till you try it.

It is much faster for a chemist to learn how to implement a recipe (such as the one above) than to master the details of a proof. And statistics often lends itself to results that can efficiently be formulated as recipes. Without a doubt there is great need for the use of statistics in business, economics, engineering, medical and social sciences, and other fields. This is the audience to which Sachs's book is addressed. These professionals are not statisticians; it seems appropriate

then to evaluate the book from their point of view.

As such, it is a book in which the statements are clear, unambiguous, and often illustrated with examples. It describes statistical concepts with great care, emphasizing aspects which are often overlooked or misinterpreted by practitioners. And (although it invites to some extent comparison with a cookbook) it offers useful insights and a great body of statistical results that are of much help to many people. The chapters on experimental design, however, could be expanded substantially and presented in a more unified way using matrix notation. The book is excellent for the audience intended. It is, however, not a book for mathematicians or statisticians. The author himself points this out in the introduction. Let me put it this way: if the null hypothesis is that the book is good, the sample I got gives results that are not significant.

I believe that books such as this one serve the useful purpose of making many statistical results accessible to researchers in other fields. It seems the answer to both questions that I initially raised is yes.

---

*Problems in Analysis*. By Bernard Gelbaum. Problems Books in Mathematics. Series Editor P. R. Halmos. Springer-Verlag, New York, 1982. vii + 228 pp.

HENRY HELSON
*Department of Mathematics, University of California, Berkeley, CA 94720*

This is the third in Springer's series of Problem Books in Mathematics, edited by P. R. Halmos, whose own book of problems about Hilbert space perhaps started a new trend. What are books of problems for? Why do publishers print them?

Books of problems are not alike. *One Hundred Problems in Elementary Mathematics* by H. Steinhaus (Pergamon Press, 1963) contains the wit and wisdom of a distinguished mathematical personality. The *Olympiad* collections from the Soviet Union have the specific purpose of identifying talent in high school students. These are two examples of problem books that are not closely related to the one under review.

In the case of Halmos' own book the answer to the questions above is not hard: the book is an authoritative account of a field of work that is currently expanding rapidly. It is useful to anyone interested in the field, from (almost) beginning graduate student to researcher. It is not only a list of problems; Halmos tells what the subject is about as he goes along, putting results in perspective and suggesting research directions. The problem format is a device that gives the author more choice as to the topics covered, and the way in which they are covered, than he would have under the constraints of an ordinary style of exposition.

The book under review is not Springer's first volume of problems in analysis. Its great predecessor is the *Aufgaben und Lehrsätze aus der Analysis* by Pólya and Szegő, 1925. That is a collection of problems *and* theorems, as the title says, and the book is still a marvelous way to study the subjects that it presents. The authors, in their preface, envisage its use as a guide for self-study. There is certainly some similarity between *Aufgaben* and the book by Halmos.

The book by Pólya and Szegő can be used in at least one very different way, however. Its subject matter begins with the elementary properties of complex numbers and infinite series, the stuff of which all mathematics was and is made. Many of the problems are excellent material for testing the ingenuity of advanced undergraduate and beginning graduate students, for informing ambitious but untried minds about themselves, for helping teachers decide who has a chance to be a mathematician. This function is particularly valuable nowadays when even mathematics majors at junior level may never have been asked to solve a problem that was not attached to a section in the text where the method of solution was expounded. Somewhere, some time, a serious student has to face the great frightening world where textbooks end. A good book of problems, like Pólya and Szegő, or Steinhaus or the *Olympiad*, can help get ready for that moment.

Gelbaum's new book is different from any of these. It contains 518 problems, all stated very briefly, with the help of elaborate notation that is clearly explained at the beginning of each section. There is a glossary of symbols, and an Index/Glossary that contains clear, concise but accurate statements of a great number of important definitions and theorems. Each problem is solved, more or less completely; the solutions occupy most of the book. Many of the problems are hard (and not only the elaborate counterexamples we are asked to find!). Quite a few look like fun, but that does not seem to have been the criterion for inclusion. On the contrary, the book is decidedly serious. This is for students who want to learn analysis.

Analysis here means what is taught in the usual graduate course on Real Variables. From that core the problems radiate into probability theory, ergodic theory, harmonic analysis and functional analysis. It is mostly analysis of the abstract kind; we find little about complex variables or differential equations. Banach spaces, measure theory, Lebesgue spaces (both abstract and on the line) are richly covered. There is no discourse, no advice, and surprisingly little reference from one problem to others. This is a problem book with a vengeance.

The author does not tell how the book can be used, so the reviewer must venture an opinion. Teachers will find an inexhaustible supply of prelim problems, and problems to assign in the course on Real Variables. That course can undoubtedly be enriched by including problems outside the syllabus proper. Students will have a splendid way to prepare for those prelims, and if they want to they can learn a lot of mathematics at the same time. The Index/Glossary is a fine way to obtain information quickly. A lecturer who does not want to follow a text closely could very well use this book in place of a text, assigning problems from it but also referring to statements and proofs of theorems. All these suggested uses are closely connected to a course on analysis. For these purposes the book is beautifully written. It is careful, very condensed but clear, and contains a great deal of fundamental material.

This is enough to ask of a book with this title, but it is not all we could hope for. Analysis has got very large, and yet today constitutes a smaller fraction of mathematics than it did a century ago. Many of our best mathematics students find analysis tedious, and who can blame them after the calculus courses we offer them? If they should have the sense to ask what analysis *is*, their instructors would be hard put to tell them. It is not that there is just one answer to such a question; but there should be at least one. In 1925 Pólya and Szegő provided their answer, and we need a modern statement of comparable conviction.

What is missing in Gelbaum's book is any intimation of what he considers important or beautiful in analysis. We can see that measure theory occupies a great deal of space, but we are not helped to know why. There are no sequences of problems leading to an area of current research. Many problems are well-known theorems, of course, but they are not grouped into a theory. At the end, as at the beginning, we do not know what the author thinks analysis is.

We welcome this admirable addition to the pedagogical literature, but fear that its use will be mainly in close connection with course work. Perhaps after all that is what the author intends.

---

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

In my note [**1**] I presented a greatly shortened and—I hope—clarified version of the standard theorem on convergence of Fourier series whose usual textbook presentation is modeled directly

on the work of Dirichlet. I want to point out that the proof in [1] is the series analogue of a corresponding argument of Ian Richards [2] for Fourier integrals. Anyone interested in basic Fourier analysis will enjoy reading Richards' elegant little paper.

**References**

1. P. R. Chernoff, Pointwise convergence of Fourier series, this MONTHLY, 87 (1980) 399–400.
2. I. Richards, On the Fourier inversion theorem for $R^1$, Proc. Amer. Math. Soc., 19 (1968) 145.

Paul R. Chernoff
Department of Mathematics
University of California
Berkeley, CA 94720

---

Editor:

Minassian (this MONTHLY, August-September 1984, p. 456) asks why referees should know the identity of authors but not vice versa. He is far from the first to ask this question. Some years ago it was asked so insistently that the American Mathematical Society instituted a policy of "blind refereeing". However, that policy has been abandoned, presumably because it was not found to help authors but did hinder referees.

The disadvantages of blind refereeing are that, first, it is often difficult for authors to hide their identities (they may, for instance, refer to their own published work). Second, it is not easy to find appropriate referees who are willing to donate their time; consequently any information that will help the referee is useful. (For instance, one is less cautious when generally careful A says "obvious" or "trivial" than when careless B says the same thing.) Third, in spite of what Minassian seems to think, open refereeing can in fact work to the disadvantage of the established author at the prestigious address. We have seen referees' reports to the effect that "this would be acceptable from a beginner but one expects better from so-and-so."

There are fairly obvious reasons why authors should not know the names of referees. Anonymity of referees avoids embarrassment; and few people are willing to act as referees if the authors are going to know who they are. Would Professor Minassian care to have to tell a personal friend, who did not show him a paper voluntarily, that he thinks it is a bad paper?

We have, between us, handled (as editors) a few thousand manuscripts, and neither of us can remember a referee's report that even hinted at rejection for nonmathematical reasons. Relatively few papers are rejected for actual errors; some are rejected for being badly written, or for having been anticipated, or for having been submitted to the wrong journal; most are rejected for not being sufficiently interesting. "Interesting" is a subjective term, and the only operational definition is that a paper is interesting if a competent referee says it is interesting. This does not mean a referee cannot support a judgment that a manuscript is of "insufficient interest", and referees ought to (and usually) do this. On the other hand, we believe the primary responsibility for demonstrating the value of a paper is the author's. The introduction or first paragraph of a manuscript is the appropriate place to explain why someone else ought to care.

Blind refereeing is almost meaningless if the editor knows who the author is. After all, the editor's values are often reflected in the choice of referee or in the phrasing of the cover letter to the referee. This is acceptable behavior, for editors are chosen because they are thought to have a certain competence, and it would be stupid to forget one's values at moments of decision. (Also one should not forget that editors are themselves authors!) The current refereeing system, being a

human endeavor, is not perfect. But we believe authors are fairly treated in the vast majority of cases.

Authors who are afraid of prejudice can rebut the reasons for rejection, or they can ask for a second referee. They can ask to have their papers refereed with the author's name omitted; most editors will honor such a request. Authors have been known to request *not* to have a particular referee; an editor would certainly take the hint.

R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201

J. J. Rotman
Department of Mathematics
University of Illinois
Urbana, IL 61801

---

Editor:

In the paper "Summing power series with polynomial coefficients" (this MONTHLY, 90 (1983) 284–285), John Klippert proves the following interesting formula:

$$f(x) = \sum_{k=0}^{p} \Delta^k a(0) x^k/(1-x)^{k+1},$$

where

$$f(x) = \sum_{n=0}^{\infty} a(n) x^n, a(n) = \sum_{k=0}^{p} c_k n^k,$$

and $\Delta$ is the well-known operator for finite differences. The formula is then illustrated by taking $a(n) = n^2 + n + 1$, which yields

$$f(x) = \sum_{n=0}^{\infty} (n^2+n+1) x^n = 2x^2/(1-x)^3 + 2x/(1-x)^2 + 1/(1-x).$$

Klippert's acceleration technique has the virtue of being accessible to a college sophomore, but it's worth noting that the same result can be obtained more easily from the well-known Taylor-like formula:

$$a(n) = \sum_{k=0}^{p} \left[\Delta^k a(0)/k!\right] n^{\underline{k}}, \quad \text{where } n^{\underline{0}} = 1, n^{\underline{k+1}} = n(n-1)^{\underline{k}}.$$

In other words, $\{n^{\underline{k}}\}_{k \in N}$ is a normal family of polynomials whose formal derivative is given by $\Delta$. Hence,

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \sum_{k=0}^{p} \left[\Delta^k a(0)/k!\right] n^{\underline{k}} x^n \\ &= \sum_{k=0}^{p} \left[\Delta^k a(0)/k!\right] x^k \sum_{n=0}^{\infty} n^{\underline{k}} x^{n-k} \end{aligned}$$

$$= \sum_{k=0}^{p} \left[\Delta^k a(0)/k!\right] x^k (1/(1-x))^{(k)}$$

and the formula is now proved.

An alternate simple proof may be given using the well-known Stirling numbers of the second kind.

The method above allows direct solutions for some practical cases, such as for the example mentioned above:

$$\begin{aligned}\sum_{n=0}^{\infty} (n^2+n+1)x^n &= \sum_{n=0}^{\infty} [n(n-1)+2n+1]x^n \\ &= x^2 \sum_{n=0}^{\infty} n(n-1)x^{n-2} + 2x \sum_{n=0}^{\infty} nx^{n-1} + \sum_{n=0}^{\infty} x^n \\ &= 2x^2/(1-x)^3 + 2x/(1-x)^2 + 1/(1-x).\end{aligned}$$

Radu Nicolescu
Computing Centre
University of Bucharest
125 Stefan Furtună Street
Bucharest, B 77116
Romania

---

## 149. MISCELLANEA

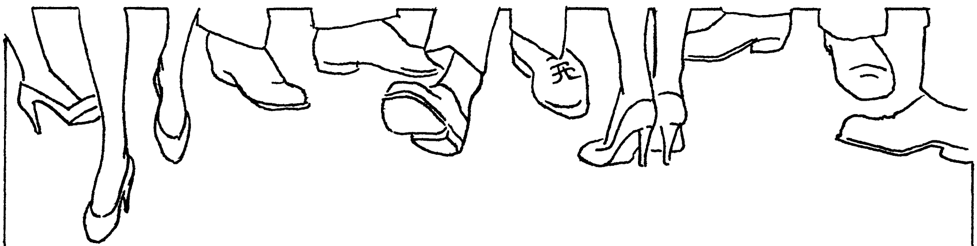
### Some More Clerihews

Saunders Mac Lane
Is eager to explain
Why mathematics is dreary
Without category theory.

Mark Kac
Had lots
Of facility
With probability.

If Max Zorn
Had never been born
We'd be in a dilemma:
What to call his lemma?

—R. P. Boas
Department of Mathematics
Northwestern University
Evanston, IL 60201



THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 6 June-July 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.




The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# RUBIK'S REVENGE: THE GROUP THEORETICAL SOLUTION

MOGENS ESROM LARSEN

*Københavns Universitets Matematiske Institut, DK-2100 København ø, Denmark*

**Introduction.** Recently Rubik and Ideal Toy have supplied us with an analog to Rubik's Cube with four slices instead of three. It is sold under the name of *Rubik's Revenge*; see Fig. 1.

The advantage is that no solution is offered besides the toy. This is however a disadvantage too, because the "Revenge" is even more difficult to solve without mathematics than the Cube and might hence disappoint a few puzzlers. I shall return to these difficulties at the end.

A general mathematician might think that this is just some more of the same. But such an abstract attitude provides no solution to the problem: fix this mess! And group theorists are used to consider groups individually; non-isomorphic groups are not much "alike".

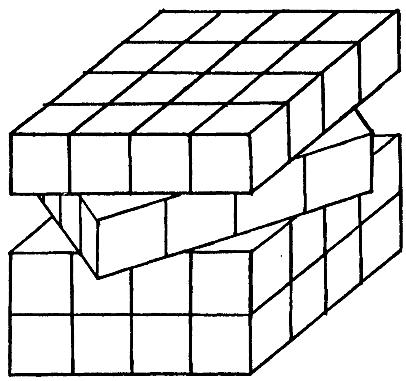


FIG. 1. Rubik's Revenge.

**Terminology.** In analogy with the Cube, see Singmaster [4], we label the six faces of the Revenge *Front*, *Back*, *Up*, *Down*, *Right*, *Left*, and we abbreviate these designations to the first letter; see Fig. 2.

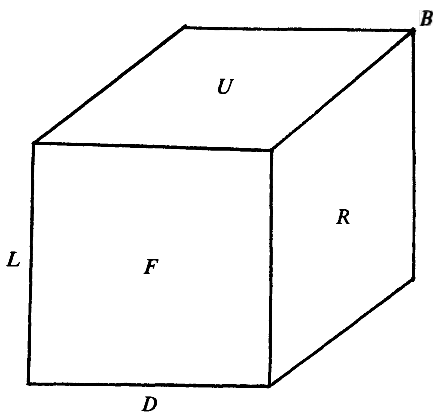



FIG. 2. Labels of faces.

*Mogens E. Larsen:* I am 42 years old, lektor (latin for "associate professor") at the department of mathematics of the University of Copenhagen, where I teach numerical analysis, applied mathematics, and occasionally complex variables, which last subject I studied at M.I.T. in the year 1969–70. Fond of games and toys I founded the Go Club of Copenhagen in 1972, and I was among the first Rubik addicts, when professor Robert Fossum from Urbana left us with a cube in the tea room of the department in the summer of 1980.

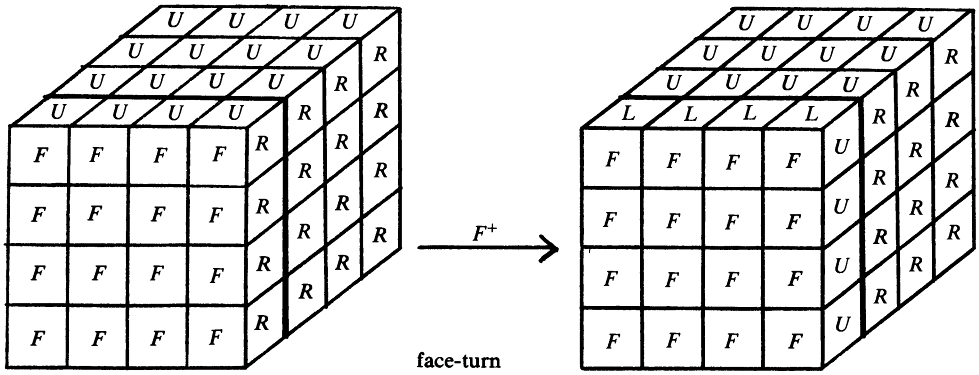



FIG. 3

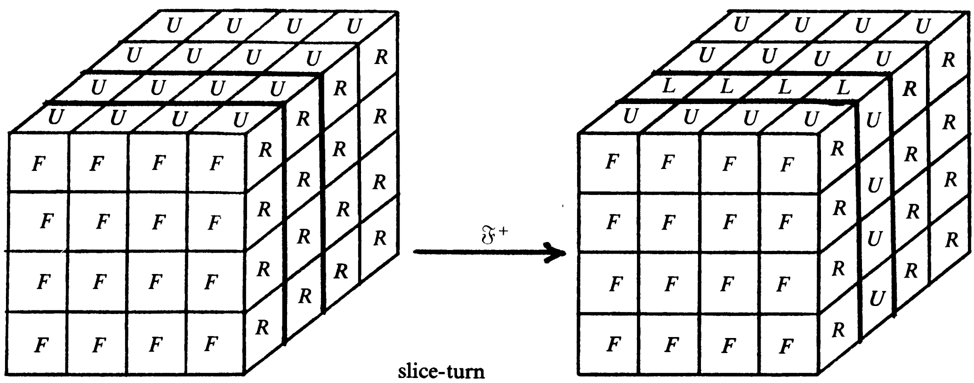



FIG. 4

A move of one of the six faces is denoted with the capital letter beginning the name of that face followed by a $+$, $-$ or 2. The "$+$" stands for turning the face 90° clockwise (seen from the outside), "$-$" for turning it 90° the other way and "2" for turning 180° (any way). Fig. 3 shows an example. This is the same as for the Cube.

For the Revenge there is also the possibility of turning the slice next to the face. We shall denote the turn of an interior slice by the gothic capital letter of the adjacent face, together with a $+$, $-$ or 2 having the same meaning as above. Fig. 4 shows an example.

An operation is denoted in the order in which the turns are supposed to be done. For example

$$F^+ L^- \mathfrak{U}^-$$

means: "Turn the front face with the clock, then turn the left face against the clock and last turn the middle upper slice against the clock."

Finally we shall call the small cube-bricks of which the big cube is built *cubinos*. These are divided in three classes known respectively as *corner-cubinos* (8 of them), *edge-cubinos* (24 of them) and *center-cubinos* (24 of them) as shown on Fig. 5.

The groups of possible operations are denoted by large capitals, while abstract groups are denoted by gothic capitals.

**The groups.** The operations form a group in a natural way. The product of two operations is merely doing these two operations one after the other in the order of writing:

$$(F^+ L^- \mathfrak{U}^-)(\mathfrak{R}^+ D^2) = F^+ L^- \mathfrak{U}^- \mathfrak{R}^+ D^2.$$

This product is obviously associative. The identity is the operation $E$ of doing "no turns".

We introduce the rules of reductions. (Symbol "$F$" may be replaced by any of the other 12 basic symbols.)

$$F^+F^- = E$$
$$F^-F^+ = E$$
$$F^+F^+ = F^2$$
$$F^-F^- = F^2$$
$$F^2F^2 = E.$$

For convenience we shall also write for the inverse

$$(F^+L^-\mathfrak{U}^-)^- = \mathfrak{U}^+L^+F^-.$$

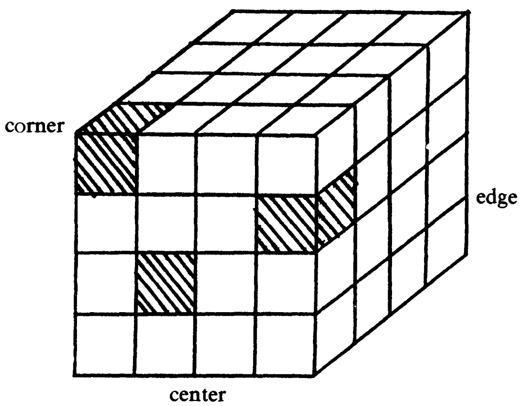



FIG. 5. Cubinos.

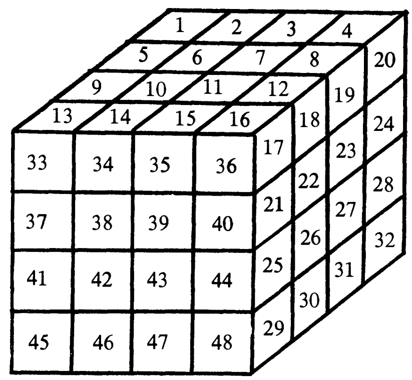



FIG. 6. Labels of cubino-faces.

Note that the inverse is obtained by putting the symbols in reverse order and changing $+-$ to $-+$. This may easily be checked by reducing as in the following example.

$$(F^+L^-\mathfrak{U}^-)(\mathfrak{U}^+L^+F^-) = E.$$

Let us denote the group of operations by $\mathfrak{G}$. This group is infinite and generated by the twelve elements of order 4:

$$F^+, B^+, U^+, D^+, R^+, L^+, \mathfrak{F}^+, \mathfrak{B}^+, \mathfrak{U}^+, \mathfrak{D}^+, \mathfrak{R}^+, \mathfrak{L}^+.$$

Let us label the 96 cubino-faces with numbers from 1 to 96; see Fig. 6. (It does not matter how they are labeled.)

Each operation corresponds to a permutation of these 96 figures, obtained by applying the operation to the labeled Revenge. Now, the product of two operations corresponds to the product of the corresponding permutations. So this correspondence is a homomorphism

$$\varphi\colon \mathfrak{G} \to \mathfrak{S}_{96},$$

where $\mathfrak{S}_n$ is the symmetric group of permutations of $n$ elements (see **[1]**, p. 54; **[3]**, p. 30).

The problem is to get a hold of $\varphi$. For example, *given* $\pi \in \mathfrak{S}_{96}$, which represents a required pattern on the Revenge, *find* an operation $X \in \mathfrak{G}$ with

$$\pi = \varphi(X),$$

means a construction of the required pattern from the solved state. Of course $\pi$ must be in the range of $\varphi$, else $X$ does not exist. Let

$$\mathbf{R} = \varphi(\mathfrak{G}) \subset \mathfrak{S}_{96}.$$

Then $\mathbf{R}$ is the group of permutations obtainable by the possible operations. The structure of $\mathbf{R}$ can be investigated in several ways.

One of the problems is to find the order of $\mathbf{R}$ (which is finite and less than $96! \simeq 10^{150}$).

Another problem is to write down in a convenient way an operation from $\varphi^{-1}(\pi)$ for any given $\pi \in \mathbf{R}$.

Let $\varepsilon$ be the identity permutation which fixes everything, and

$$\mathfrak{E} = \varphi^{-1}(\varepsilon) \subset \mathfrak{G}$$

the kernel of $\varphi$. Then $\mathfrak{E}$ is the group of operations leaving all cubino-faces unmoved. Thus we have a factor group

$$\mathbf{R} \simeq \mathfrak{G}/\mathfrak{E}.$$

It is not possible by any operation to turn any cubino-face around itself on the place where it is, because both types of operations, $F^+$ and $\mathfrak{F}^+$, preserve orientation, that is, the corner nearest the center remains nearest the center; see Fig. 7.

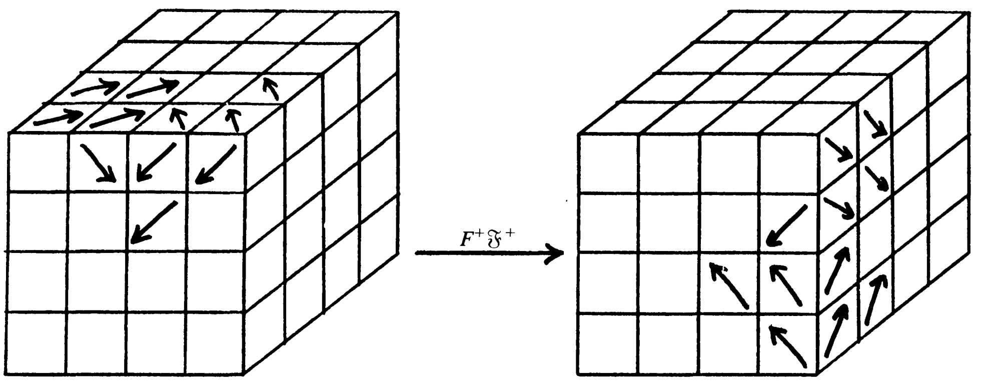



FIG. 7. Preservation of cubino-face orientation.

But it is possible to permute the four equally coloured center-cubinos among themselves. This means that the order of $\mathbf{R}$ is bigger than the number of visible patterns. Therefore, let $\mathbf{C}_0$ denote the subgroup of $\mathbf{R}$ permuting each of the six classes of four equal center-cubinos among themselves (i.e., $\mathbf{C}_0$ consists of the invisible operations). This group $\mathbf{C}_0$ is not normal in $\mathbf{R}$, but nevertheless, the number of patterns of the Revenge is the number of cosets to $\mathbf{C}_0$ in $\mathbf{R}$. (This is to say, two patterns in $\mathbf{R}$ look the same, if one is obtained from the other by permutation of some equal center-cubinos of like color.)

**The structure of the group R.** Each operation creates a permutation of the 56 cubinos and sometimes a turn of some of them too. But is it not possible to move any cubino to an arbitrary prescribed position.

The 8 corner-cubinos are permuted among themselves and if possible turned individually.

The 24 edge-cubinos are permuted among themselves and if possible flipped individually (we shall see that this flip is actually impossible).

The 24 center-cubinos are permuted among themselves, but as explained above, they cannot be rotated individually.

This means that each operation creates three simultaneous permutations of the three kinds mentioned above together with some turns and flippings. We shall now examine each of these kinds in more detail.

*The corner-cubinos*

First we consider the corners ignoring edges and centers. (We can imagine the Revenge with

edges and centers painted all black.) Then the most we can hope to do with the corner-cubinos is to obtain any permutation of them from $\mathfrak{S}_8$ together with any 8-tuple of turns from $\mathfrak{Z}_3$ (the cyclic group of order 3). This means a total of $3^8 8!$ patterns.

Group-theoretically we consider these patterns as forming a group of permutations, the so-called wreath product of $\mathfrak{Z}_3$ by $\mathfrak{S}_8$, written

$$\mathfrak{Z}_3 \wr \mathfrak{S}_8$$

(see [1], p. 81; [4], p. 59).

Intuitively the group $\mathfrak{Z}_3 \wr \mathfrak{S}_8$ contains a subgroup leaving all corner-cubinos on their places, but turning them individually. This subgroup is normal and isomorphic to $\mathfrak{Z}_3^8$. The corresponding factorgroup consists of the permutations of the 8 corner-cubinos ignoring their possible turns, that is

$$(\mathfrak{Z}_3 \wr \mathfrak{S}_8)/\mathfrak{Z}_3^8 \simeq \mathfrak{S}_8.$$

The possible operations create a subgroup $\mathfrak{H}$ of $\mathfrak{Z}_3 \wr \mathfrak{S}_8$. This is the same as it is for the standard Cube, because the corner-cubinos are not affected by the slice-turns. From the analysis of the Cube (see [2], p. 50; [4], p. 17) we know that $\mathfrak{H}$ has index 3 in $\mathfrak{Z}_3 \wr \mathfrak{S}_8$. Indeed, the sum of all turns of corner-cubinos must be divisible by 3.

Next we consider the subgroup $\mathbf{H}$ of $\mathbf{R}$ consisting of those permutations obtainable by operations fixing edge-cubinos and center-cubinos. The group $\mathbf{H}$ is isomorphic to a subgroup of $\mathfrak{H}$. The group $\mathbf{H}$ is as above the same for the Revenge as for the Cube and hence of index 2 in $\mathfrak{H}$. It contains the subgroup of turns without permutations, $\mathbf{T}$,

$$\mathbf{T} \simeq \mathfrak{Z}_3^7,$$

and the corresponding factorgroup is isomorphic to $\mathfrak{A}_8$, the alternating group of even permutations of 8 elements (see [1], p. 59; [3], p. 32),

$$\mathbf{H}/\mathbf{T} \simeq \mathfrak{A}_8.$$

Hence the order of $\mathbf{H}$ is $3^7 \cdot \frac{1}{2} \cdot 8!$

The fact that $\mathbf{H}$ has index 2 in $\mathfrak{H}$ means that we can only create an odd permutation of corner-cubinos at the expense of some odd permutation of either edge-cubinos or center-cubinos. We shall return to this problem in the following section.

*The center-cubinos*

First we consider the center-cubinos ignoring corners and edges. Then the most we can hope to do with the center-cubinos is to obtain any permutation of them from $\mathfrak{S}_{24}$, because they cannot be rotated, as already mentioned.

The possible operations create a subgroup $\mathfrak{C}$ of $\mathfrak{S}_{24}$. We shall compute $\mathfrak{C}$ below.

Next we consider the subgroup $\mathbf{C}$ of $\mathbf{R}$ consisting of those permutations obtainable by operations fixing edge-cubinos and corner-cubinos. The group $\mathbf{C}$ is isomorphic to a subgroup of $\mathfrak{C}$, let us write

$$\mathbf{C} \subset \mathfrak{C} \subset \mathfrak{S}_{24}.$$

It turns out that $\mathbf{C}$ is isomorphic to the alternating group of even permutations of 24 elements, and that $\mathfrak{C}$ is isomorphic to the symmetric group itself.

Before computing these groups we shall recall a couple of notations from group theory.

The *commutator* of two group elements $X$ and $Y$ is defined as

$$(\textit{commutator}) \qquad [X, Y] = XYX^{-}Y^{-}$$

(see [1], p. 138; [2], p. 33; [3], p. 48; [4], p. 17).

REMARK 1. Two group elements $X$ and $Y$ commute exactly when the commutator of them is the neutral element, $[X, Y] = E$.

REMARK 2. The inverse of a commutator is the "reverse"

$$[X,Y]^{-}=[Y,X].$$

The *conjugate* of a group element $X$ *under* the group element $Y$ is

$$(\textit{conjugate}) \qquad YXY^{-},$$

(see [1], p. 13; [2], p. 31; [4], p. 13).

REMARK 3. On the Cube and the Revenge we consider the conjugate of a "nice" operation $X$ as follows. The operation $X$ is for example a cycle of 3 cubinos. The operation $Y$ moves 3 cubinos which we want to cycle to the locations where $X$ operate. The irrelevant mess $Y$ might create is cleared up by $Y^{-}$ at the same time as the 3 wanted cubinos are returned to the original 3 places, but cycled. (This fact is reckoned in mathematics as the theorem that conjugate permutations have the same structure of cycles; see [1], p. 54.)

REMARK 4. It is fairly easy to see that there is a possible permutation in **R** which moves 3 distinct center-cubinos to any stipulated locations (no assumption is made about other cubinos). The same is independently true about edge-cubinos.

We shall now return to the computation of the center-groups.

THEOREM 1. $\mathbf{C} \simeq \mathfrak{A}_{24}$ *and* $\mathfrak{C} \simeq \mathfrak{S}_{24}$.

*Proof*. The group **C** contains the 3-cycle (1), see Fig. 8.

$$(1) \qquad [[\mathfrak{F}^{+},\mathfrak{D}^{+}],U^{-}] \in \mathbf{C}.$$

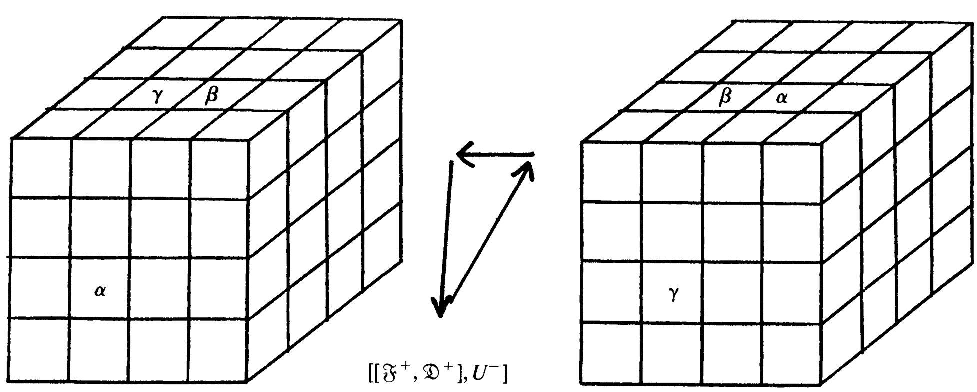



FIG. 8. A 3-cycle of center-cubinos.

It is obvious from Remarks 3 and 4 that we can obtain *any* 3-cycle of center-cubinos by conjugation of this one. Hence **C** contains all the 3-cycles of center-cubinos.

Any alternating group is generated by the set of 3-cycles (see [1], p. 61; [3], p. 33). Hence we must have the inclusion

$$\mathfrak{A}_{24} \subset \mathbf{C} \subset \mathfrak{C} \subset \mathfrak{S}_{24}.$$

Now, $F^{+} \in \mathfrak{C}$ is a cycle of length 4 and hence an odd permutation. So $\mathfrak{C}$ is greater than $\mathfrak{A}_{24}$ and we must have

$$\mathfrak{C} \simeq \mathfrak{S}_{24}.$$

On the other hand, $F^{+} \in \mathbf{R}$ consists of 4 cycles of length 4 and hence is even, see Fig. 9. Indeed, it contains one 4-cycle of corner-cubinos, one 4-cycle of center-cubinos and two 4-cycles of edge-cubinos. Hence it can only give an odd permutation of center-cubinos simultaneously with an odd permutation of corner-cubinos. So, when corner-cubinos are kept fixed (the even permutation of nothing), then the center-cubinos must be permuted by an even permutation.

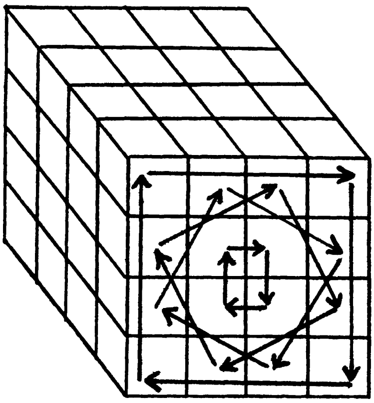


FIG. 9. The 4 4-cycles of $F^+$.

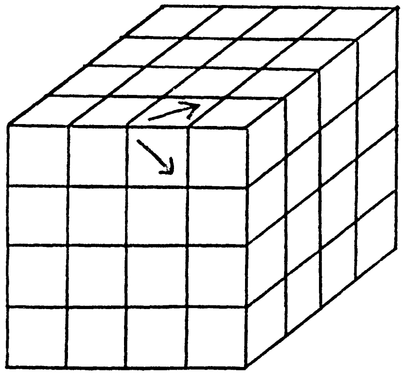


FIG. 10. Preservation of orientation.

And $\mathfrak{F}^+ \in \mathbf{R}$ consists of 3 cycles of length 4, so this is odd. But one of them is a 4-cycle of edge-cubinos and the remaining two are 4-cycles of center-cubinos. So, considered as a permutation of center-cubinos $\mathfrak{F}^+$ is even.

Conclusion. No operation in $\mathbf{C}$ can be an odd permutation. Hence

$$\mathbf{C} \simeq \mathfrak{A}_{24}. \quad \square$$

The order of $\mathfrak{C}$ then is 24!, and for $\mathbf{C}$ it is $\frac{1}{2} \cdot 24!$.

The fact that $\mathbf{C}$ has index 2 in $\mathfrak{C}$ means that we are only able to create an odd permutation of center-cubinos at the expense of some other odd permutation. From the proof it follows that the other odd permutation must be a permutation of the corner-cubinos.

The last fact may be explained further. Consider the group $\mathfrak{P}$ of permutations of corner-cubinos and center-cubinos while ignoring the edge-cubinos. The group $\mathfrak{P}$ must be isomorphic to a subgroup of $\mathfrak{H} \times \mathfrak{C}$. The meaning of the restriction above is then that a pair of permutations $(X, Y) \in \mathfrak{H} \times \mathfrak{C}$ belongs to $\mathfrak{P}$ if and only if the permutations $X \in \mathfrak{H}$ and $Y \in \mathfrak{C}$ have the same parity. This means that $\mathfrak{P}$ is the subgroup of $\mathfrak{H} \times \mathfrak{C}$ of index 2 consisting of all operations that are even as permutations of these 32 cubinos.

Still unsettled is the question of the group $\mathbf{P}$ consisting of those operations from $\mathfrak{P}$ which fix all edge-cubinos. Obviously they include all pairs from $\mathbf{H} \times \mathbf{C}$. We can thus say

$$\mathbf{H} \times \mathbf{C} \subset \mathbf{P} \subset \mathfrak{P},$$

and as $\mathbf{H} \times \mathbf{C}$ has index 2 in $\mathfrak{P}$, we know now that $\mathbf{P}$ is one or the other.

When we have computed the edge-groups, we shall be able to show that

$$\mathbf{P} \simeq \mathfrak{P}.$$

*The edge-cubinos*

First we consider the edge-cubinos ignoring corners and centers. Then the most we can hope to do with the edge-cubinos is to obtain any permutation of them from $\mathfrak{S}_{24}$ together with any 24-tuple of flips from $\mathfrak{Z}_2$. It turns out that edge-cubinos cannot be flipped, so it is enough to consider the subgroups of $\mathfrak{S}_{24}$ and we can avoid the wreath product.

THEOREM 2. *No operation flips any edge-cubino.*

*Proof.* If an edge-cubino is flipped, then the orientation from Fig. 7 must look like Fig. 10. But this is not obtainable, because the cubino-faces then do not preserve orientation. $\square$

REMARK. It is also true that if an edge-cubino is physically removed, the hidden "foot" will be revealed to be asymmetric.

This means that for any given location and any given edge-cubino, it can only be oriented in one way. Hence we can consider the group of operations on the edges as a permutation group $\mathfrak{K}$ of the 24 edge-cubinos, i.e., $\mathfrak{K} \subset \mathfrak{S}_{24}$.

Next we consider the subgroup **K** of **R** of operations on the edge-cubinos fixing the center-cubinos and the corner-cubinos.

We have from Theorem 2, that

$$\mathbf{K} \subset \mathfrak{K} \subset \mathfrak{S}_{24}.$$

They all turn out to be equal.

THEOREM 3.

$$\mathbf{K} \simeq \mathfrak{K} \simeq \mathfrak{S}_{24}.$$

*Proof*. At first we prove that

$$\mathfrak{A}_{24} \subset \mathbf{K}.$$

The group **K** contains the 3-cycle (2); see Fig. 11.

(2) $$[\mathfrak{L}^-,[L^+,U^-]] \in \mathbf{K}.$$

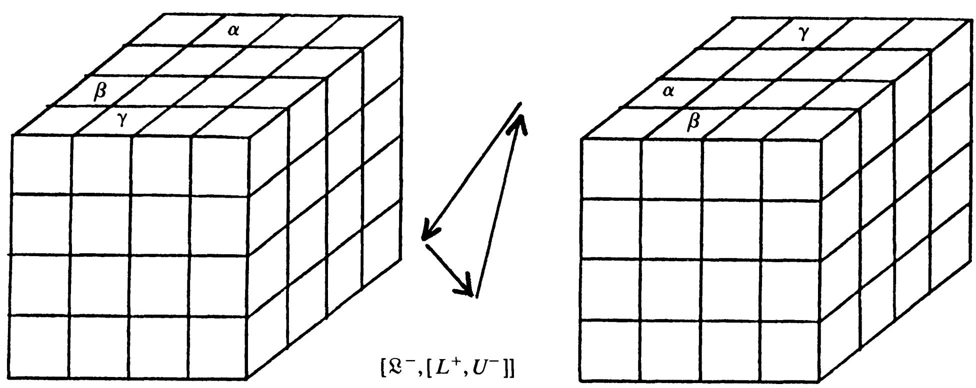



FIG. 11. A 3-cycle of edge-cubinos.

It is obvious from Remarks 3 and 4 that we can obtain *any* 3-cycle of edge-cubinos by conjugation of this one. Hence **K** contains all 3-cycles of edge-cubinos and thereby $\mathfrak{A}_{24}$.

It is obvious that $\mathfrak{K}$ contains an odd permutation, $\mathfrak{F}^+ \in \mathfrak{K}$ is a 4-cycle of edge-cubinos and odd.

But $\mathfrak{F}^+$ indicates the existence of a 4-cycle in **K** too. $\mathfrak{F}^+$ consists of a 4-cycle of edge-cubinos together with two 4-cycles of center-cubinos. A permutation of center-cubinos made of two 4-cycles is even and hence belongs to **C** according to Theorem 1. This means that we can rearrange the center-cubinos with the help of a permutation from **C**, that is a permutation fixing all other cubinos. Doing this we are left with exactly one 4-cycle of edge-cubinos, which is the wanted odd permutation in **K**. So **K** is bigger than $\mathfrak{A}_{24}$ and hence equal to $\mathfrak{S}_{24}$. □

The orders of **K** and $\mathfrak{K}$ are both 24!.

The fact that $\mathbf{K} \simeq \mathfrak{K}$ $(\simeq \mathfrak{S}_{24})$ means that any permutation of edge-cubinos can be done without disturbing the others. This also means that any permutation in $\mathfrak{P}$ can be done without changing the edge-cubinos. So from this follows, that

$$\mathbf{P} \simeq \mathfrak{P}.$$

*The structure of* **R**

The result in Theorem 2 says about **R**, that

$$\mathbf{R} \simeq \mathbf{K} \times \mathbf{P}.$$

Any possible permutation can be executed independently on the edge-cubinos and on the other cubinos. And from Theorem 1 we have

$$\mathbf{H} \times \mathbf{C} \subset \mathbf{P} \subset \mathfrak{H} \times \mathfrak{C},$$

where $\mathbf{P}$ consists of the even permutations. This is to say, that of the 3 subgroups containing $\mathbf{H} \times \mathbf{C}$ and of index 2 in $\mathfrak{H} \times \mathfrak{C}$, the group $\mathbf{P}$ is the one different from $\mathbf{H} \times \mathfrak{C}$ and $\mathfrak{H} \times \mathbf{C}$. See Fig. 12.

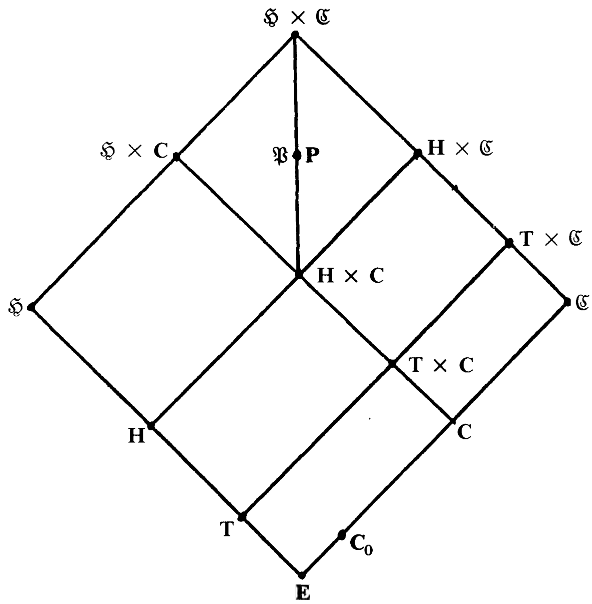



FIG. 12. Diagram of center- and corner-groups.

The order of $\mathbf{R}$ then is

$$\mathrm{order}(\mathbf{R}) = \mathrm{order}(\mathbf{K}) \cdot \mathrm{order}(\mathbf{P}),$$

and

$$\mathrm{order}(\mathbf{P}) = 2 \cdot \mathrm{order}(\mathbf{H}) \cdot \mathrm{order}(\mathbf{C});$$

all together

$$\mathrm{order}(\mathbf{R}) = 24! \cdot 2 \cdot 3^7 \cdot \frac{1}{2} \cdot 8! \cdot \frac{1}{2} \cdot 24!.$$

To compute the number of patterns recall the group $\mathbf{C}_0$ of invisible permutations. As $\mathbf{C}_0$ is a subgroup of $\mathbf{C}$, it must consist of the even permutations of each of the six 4-tuples by itself. There are $4!^6$ permutations of 6 sets of 4 elements each, and of these, half are even. The order of $\mathbf{C}_0$ hence is

$$\frac{1}{2} \cdot 24^6.$$

The number of patterns of visible difference hence is

$$\frac{3^7 \cdot 8! \cdot 24!^2}{24^6} = 177\ 62872\ 41975\ 57644\ 87697\ 82553\ 87965\ 78406\ 40000\ 00000.$$

(This number includes a factor 24 from the group of physical turns of the cube (the *hexahedron* group; see [3], p. 37).)

**How to solve a mess?** The principle of solution may be based on a descending series of subgroups, each normal in the preceding one with a nice factor group. The use of the series shall be explained below. Let us define $\mathbf{K}' \simeq \mathfrak{A}_{24}$ as the subgroup of $\mathbf{K}$ of even permutations, and $\mathbf{E}$ as the group of one element, the unit. We shall then give an example of such a series (with factor groups below):

$$\begin{array}{ccccccccccccc} \mathbf{R} & \supset & \mathbf{K}\times\mathbf{H}\times\mathbf{C} & \supset & \mathbf{K}\times\mathbf{T}\times\mathbf{C} & \supset & \mathbf{K}\times\mathbf{E}\times\mathbf{C} & \supset & \mathbf{K}'\times\mathbf{E}\times\mathbf{C} & \supset & \mathbf{E}\times\mathbf{E}\times\mathbf{C} & \supset & \mathbf{E}\times\mathbf{E}\times\mathbf{E} \\ & \mathfrak{Z}_2 & & \mathfrak{A}_8 & & \mathbf{T}\simeq\mathfrak{Z}_3^7 & & \mathfrak{Z}_2 & & \mathfrak{A}_{24} & & \mathfrak{A}_{24} & \end{array}$$

We consider the pattern (the mess) of the Revenge as a permutation obtained from the nice start position with one-coloured faces by some operation. This permutation belongs to $\mathbf{R}$, and we want to find a way to write it as a series of the twelve generators of $\mathbf{R}$, $F^+$, $\mathfrak{F}^+$, etc.

If the permutation of corner-cubinos is odd, then we start with any face-turn, for example $F^+$. This is to be considered as a representation of the non-unit of the first factor group $\mathfrak{Z}_2$. Thus we reduce the problem to a permutation in $\mathbf{K}\times\mathbf{H}\times\mathbf{C}$.

Then we use the generators of $\mathbf{H}/\mathbf{T} \simeq \mathfrak{A}_8$, the 3-cycles of corners known from the Cube; see [2], p. 32; [4], p. 44, e.g.,

$$[F^-, U^+B^+U^-],$$

to arrange the corner-cubinos on right places relative to each other. This means we have reduced the problem to a permutation in the group $\mathbf{K}\times\mathbf{T}\times\mathbf{C}$.

Now the generators of $\mathbf{T} \simeq \mathfrak{Z}_3^7$ turn the corners in the right directions. They are also obtainable from the Cube; see [2], p. 38; [4], p. 44, e.g.,

$$\left[[F^+, D^+]^2, U^+\right].$$

This reduces the problem to a permutation in the group $\mathbf{K}\times\mathbf{E}\times\mathbf{C}$, and having arranged the corners we rejoice that permutations from this group keep corners fixed.

If the necessary permutation of edge-cubinos is odd, then we need a generator of $\mathbf{K}/\mathbf{K}' \simeq \mathfrak{Z}_2$; i.e., any odd permutation, e.g.,

$$\mathfrak{F}^+.$$

After this reduction to the group $\mathbf{K}'\times\mathbf{E}\times\mathbf{C}$, we can replace the edge-cubinos completely by help of formula (2) and its conjugates. This replacement reduces our permutation to the group $\mathbf{E}\times\mathbf{E}\times\mathbf{C}$.

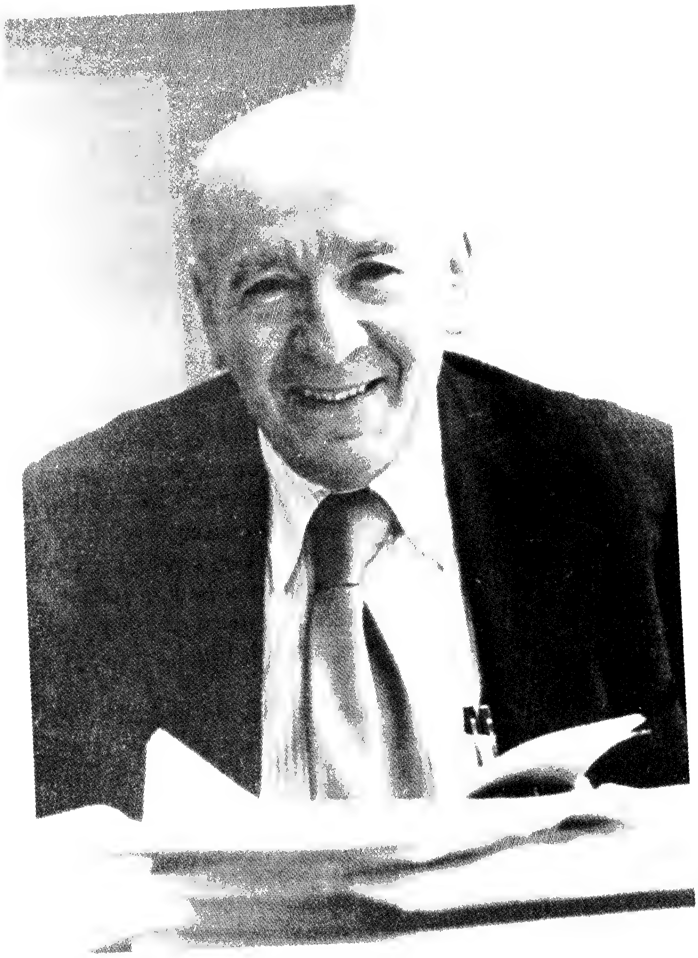
This group is generated by the formula (1) and its conjugates. Using this we don't need to reach $\mathbf{E}$; it is enough to reduce to any permutation from $\mathbf{C}_0$ (the difference is invisible).

The descending series used here is vivid graphically, because more and more cubinos fall into place. The lower subgroups tend to have more complicated operations. A series can be constructed on the opposite principle, where the last stages use simpler operations. Watching the result is like magic. Nothing seems to be happening until the last few moments, when PRESTO! everything falls into place.

The problems in the case of the Revenge are due to the two factor groups $\mathfrak{Z}_2$. They are so easy to overlook, but to postpone these two "corrections" might double the work. Besides, they are hard to invent by oneself —a mathematical benefit from having heard about odd and even!

## References

**0.** Christoph Bandelow, Inside Rubik's Cube™ and Beyond, Basel, 1982.
**1.** Marshall Hall, The Theory of Groups, New York, 1959.
**2.** Mogens Esrom Larsen, Rubiks terning, København, 1981.
**3.** ______, Gruppeteori, København, 1982.
**4.** David Singmaster, Notes on Rubik's 'Magic Cube', 5th ed., London, 1980.



A singular man who modestly disclaimed being essential. (See p. 406.)

$|\arg f(z)| < \alpha\frac{\pi}{2}$ and $f(0) = 1$, with $\alpha > 1$. In each the set of support points is the same as $EH\mathscr{F}$ [5], [13].

## References


**1.** Y. Abu-Muhanna and T. H. MacGregor, Extreme points of families of analytic functions subordinate to convex mappings, Math. Z., 176 (1981) 511–519.

**2.** L. V. Ahlfors, Complex Analysis, 3rd edition, McGraw-Hill, New York, 1979.

**3.** R. V. Benson, Euclidean Geometry and Convexity, McGraw-Hill, New York, 1966.

**4.** D. A. Brannan, J. G. Clunie, and W. E. Kirwan, On the coefficient problem for functions of bounded boundary rotation, Ann. Acad. Sci. Fenn. Ser. A I Math., 523 (1973) 18 pp.

**5.** L. Brickman, T. H. MacGregor, and D. R. Wilken, Convex hulls of some classical families of univalent functions, Trans. Amer. Math. Soc., 156 (1971) 91–1

**6.** C. Cochrane and T. H. MacGregor échet differentiable functionals and support points of families of analytic functions, Trans. Amer. Math. Soc., 236 (1978) 75–92.

**7.** N. Dunford and J. T. Schwartz, Linear Operators. I: General Theory, Pure and Appl. Math., vol. 7, Interscience, New York, 1958.

**8.** P. L. Duren, Theory of $H^p$ Spaces, Academic Press, New York, 1970.

**9.** ______, Univalent Functions, Springer-Verlag, New York, 1983.

**10.** H. G. Eggleston, Convexity, Cambridge Univ. Press, Cambridge, 1963.

**11.** A. W. Goodman, The valence of sums and products, Canad. J. Math., 20 (1968) 1173–1177.

**12.** D. J. Hallenbeck and T. H. MacGregor, Subordination and extreme-point theory, Pacific J. Math., 50 (1974) 455–468.

**13.** ______, Support points of families of analytic functions described by subordination, Trans. Amer. Math. Soc., 278 (1983) 523–546.

**14.** ______, Linear Problems and Convexity Techniques in Geometric Function Theory, Pitman Publishing Co., Boston, 1984.

**15.** D. H. Hamilton, The extreme points of $\Sigma$, Proc. Amer. Math. Soc., 85 (1982) 393–396.

**16.** G. Herglotz, Über Potenzreihen mit positivem, reelem Teil im Einheitskreis, S.-B. Sächs Akad. Wiss. Leipzig Math.-Natur. Kl., 63 (1911) 501–511.

**17.** T. H. MacGregor, Applications of extreme-point theory to univalent functions, Michigan Math. J., 19 (1972) 361–376.

**18.** ______, Hull subordination and extremal problems for starlike and spirallike mappings, Trans. Amer. Math. Soc., 183 (1973) 499–510.

**19.** ______, Variability regions for families of univalent functions, Indiana Univ. Math. J., 31 (1982) 689–699.

**20.** J. G. Milcetich, On the extreme points of some sets of analytic functions, Proc. Amer. Math Soc., 45 (1974) 223–228.

**21.** R. R. Phelps, Lectures on Choquet's Theorem, Van Nostrand, New York, 1966.

**22.** G. Pólya and G. Szegő, Problems and Theorems in Analysis (2 vols.), Springer-Verlag, Berlin, 1972, 1976.

**23.** Chr. Pommerenke, Univalent Functions, Vandenheock and Ruprecht, Göttingen, 1975.

**24.** P. Porcelli, Linear Spaces of Analytic Functions, Rand McNally, Chicago, 1966.

**25.** F. Riesz, Sur certains systèmes singuliers d'équations intégrales, Ann. École Norm. Sup., 28 (1911) 33–62.

**26.** R. M. Robinson, Univalent majorants, Trans. Amer. Math. Soc., 61 (1947) 1–35.

**27.** St. Ruscheweyh, Nichtlineare Extremalprobleme für holomorphe Stieltjesintegrale, Math. Z., 142 (1975) 19–23.

**28.** G. Schober, Univalent Functions-Selected Topics, Lecture Notes in Mathematics no. 478, Springer-Verlag, Berlin, 1975.

**29.** H. Silverman, Complex Variables, Houghton Mifflin, Boston, 1975.

**30.** G. Springer, Extreme Punkte der konvexen Hülle schlichte Funktionen, Math. Ann., 129 (1955) 230–232.

**31.** M. Tsuji, Potential Theory in Modern Function Theory, Maruzen Co., Tokyo, 1975.

**32.** Nguyen Xuan Uy, Removable sets of analytic functions satisfying a Lipschitz condition, Ark. Mat., 17 (1979) 19–27.

**33.** F. A. Valentine, Convex Sets, McGraw-Hill, New York, 1964.


---

## ANSWER TO PHOTO ON PAGE 391

Mark Kac (1914–1984) enjoyed describing himself as only a simple Pole.

# LINEAR METHODS IN GEOMETRIC FUNCTION THEORY

T. H. MacGREGOR
*Department of Mathematics and Statistics, SUNY at Albany, Albany, NY 12222*

**1. Introduction.** This paper describes results in classical complex analysis in which convexity is the main theme. More generally, linear and topological ideas are central to the presentation. The emphasis is on finding the extreme points of families of analytic functions arising in geometric function theory and applying this information to the solution of extremal problems. In particular, such questions are discussed in the context of starlike and convex mappings of the open unit disk and other families defined geometrically.

Some results mentioned are well known and others were obtained recently, a few only in the last couple of years. The approach represented by this development provides a way of deriving several established results from a coherent point of view based on a few general abstract principles, the most critical of which is the Krein-Milman theorem. This approach also leads to the solution of more difficult problems. The presentation here focuses on certain families and problems where there are complete and striking results.

Several ideas used in this paper are geometric. Some only involve convexity facts about the set of complex numbers. A useful reference for such ideas in $R^n$, Euclidean $n$-space, is **[10]**. Usually the objects to be studied here are sets of analytic functions, and thus certain concepts require a more abstract setting. A suitable one is that of a locally convex linear topological space. Chapter 5 of the book **[7]** contains the facts needed about such spaces. Two additional useful references about convexity are **[3]** and **[33]**, which are at a more elementary level.

Linear methods are well established in areas of complex analysis, such as in the theory of $H^p$ spaces **[8]**. It was much more recently that such techniques had an influence on research in the theory of univalent functions. A number of results about linear problems and univalent functions are contained in **[28]**, and a forthcoming monograph **[14]** has a similar orientation. This article represents a survey of these developments.

The author began writing this paper in preparation for an invited address given at the 1983 summer meeting of the Mathematical Association of America held in Albany, New York, on August 7–11. The title of the talk was "Convexity ideas in geometric function theory." The author dedicates this paper to Vincent Cowling, who encouraged the author to give that talk and has been supportive of the author's professional work for several years.

**2. Linear, topological structure of the space of analytic functions.** Let $\mathscr{A}$ denote the set of functions that are analytic in the open unit disk $\Delta = \{z\colon\ |z| < 1\}$. The set $\mathscr{A}$ is a vector space with respect to the usual addition of functions and multiplication of functions by complex numbers. A topology for $\mathscr{A}$ is described below; it corresponds to convergence which is uniform on compact subsets of $\Delta$. This topology is chosen because it so intimately relates to classical extremal problems for $\mathscr{A}$. An important theorem of Weierstrass asserts that if $\{f_n\}$ is a sequence of functions analytic on a region $D$ that converges uniformly on each compact subset of $D$, then $f = \lim f_n$ is also analytic in $D$ **[2**, p. 176]. Moreover, $\{f_n'\}$ converges uniformly to $f'$ on each compact subset of $D$.

The topology for $\mathscr{A}$ is defined as follows. Let $M(r, f) = \max\{|f(z)|: |z| \leqslant r\}$ when $f \in \mathscr{A}$ and $0 < r < 1$. Let $\{r_n\}$ be an increasing sequence of real numbers such that $0 < r_n < 1$

---

Thomas H. MacGregor graduated from Lafayette College in 1954 and received the Ph.D. in mathematics from the University of Pennsylvania in 1961. His dissertation advisor was O. H. Ališbah and his graduate education included classes given by B. Epstein, E. Grosswald, H. Rademacher, and I. J. Schoenberg. MacGregor's research is in complex analysis in the field of geometric function theory. He was a faculty member at Lafayette College from 1959 to 1967 and at SUNY at Albany since 1967. MacGregor served as chairman of the Mathematics department at Albany from 1975 to 1978. He is an ardent fisherman and gardener. When successful, these activities are appreciated by his wife Billie and son Tom.

$(n = 1, 2, \ldots)$ and $r_n \to 1$. If $f, g \in \mathscr{A}$ and $d_n(f, g) = M(r_n, f - g)$, then $d$ is defined by

$$d(f,g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(f,g)}{1 + d_n(f,g)}. \tag{1}$$

It is not difficult to show that $d$ is a metric on $\mathscr{A}$ and the topology given by $d$ is equivalent to convergence that is uniform on each compact subset of $\Delta$ [**2**, p. 221]. As usual for a metric space, $f_n \to f$ means $d(f_n, f) \to 0$. More generally the vector space $\mathscr{A}(D)$ of functions analytic on a region $D$ can be described the same way. In particular, the metric for $\mathscr{A}(D)$ is obtained by replacing the sets $\{z : |z| \leqslant r_n\}$ by an increasing sequence of compact sets whose union is $D$.

A fundamental theorem of P. Montel is especially useful in determining topological properties of subsets of $\mathscr{A}(D)$. If $\mathscr{F} \subset \mathscr{A}(D)$, then $\mathscr{F}$ is said to be locally uniformly bounded provided that the following condition holds: if $C$ is a compact subset of $D$, then there is a positive real number $M$ such that $|f(z)| \leqslant M$ when $f \in \mathscr{F}$ and $z \in C$. Montel's theorem states that each sequence $\{f_n\}$ of functions belonging to a locally uniformly bounded family has a convergent subsequence [**2**, p. 224]. Montel's theorem is a generalization of the Bolzano-Weierstrass theorem, which asserts that every bounded sequence of complex numbers has a convergent subsequence. In fact, the proof of Montel's theorem involves a repeated application of the Bolzano-Weierstrass theorem. Since $\mathscr{A}(D)$ is a metric space, Montel's theorem leads to the following characterization of compact sets: a subset $\mathscr{F}$ of $\mathscr{A}(D)$ is compact if and only if $\mathscr{F}$ is locally uniformly bounded and whenever $f_n \in \mathscr{F}$ and $f_n \to f$ then $f \in \mathscr{F}$.

The space $\mathscr{A}(D)$ also is locally convex. In particular for $\mathscr{A}$, $\{f : |f(z)| < M \text{ when } |z| < r\}$ is a convex neighborhood of 0 for each $M$ and $r$ such that $M > 0$ and $0 < r < 1$. In summary, $\mathscr{A}(D)$ has the properties of a locally convex linear topological space. Thus, the ideas and results from that general theory are applicable, and they shall be presented here generally in a form that applies to $\mathscr{A}$.

DEFINITION. If $\mathscr{F} \subset \mathscr{A}$, then $f$ is called an extreme point of $\mathscr{F}$ provided that $f \in \mathscr{F}$ and the relations $f = tg + (1 - t)h$, $g \in \mathscr{F}$, $h \in \mathscr{F}$, and $0 < t < 1$, imply that $g = h$.

In other words, $f$ is not a proper convex combination of two distinct elements of $\mathscr{F}$. Let $E\mathscr{F}$ denote the set of extreme points of $\mathscr{F}$. A general problem discussed here is finding $E\mathscr{F}$ for certain subsets of $\mathscr{A}$.

### 3. Functions having a positive real part.

DEFINITION. Let $\mathscr{P}$ denote the subset of $\mathscr{A}$ consisting of functions $p$ such that $\operatorname{Re} p(z) > 0$ $(|z| < 1)$ and $p(0) = 1$.

An example of a member of $\mathscr{P}$ is given by $P(z) = (1 + z)/(1 - z)$, which maps $\Delta$ one-to-one onto the half plane $\{w : \operatorname{Re} w > 0\}$. Clearly, $\mathscr{P}$ is a convex subset of $\mathscr{A}$. Schwarz's lemma [**2**, p. 135] is used here and later on. It asserts that if $\phi$ is analytic in $\Delta$, $|\phi(z)| < 1$ for $|z| < 1$ and $\phi(0) = 0$, then $|\phi(z)| \leqslant |z|$. If $p \in \mathscr{P}$, then $\phi = P^{-1} \circ p$ satisfies Schwarz's lemma. This implies that if $p \in \mathscr{P}$, then $|p(z)| \leqslant (1 + |z|)/(1 - |z|)$. Therefore $\mathscr{P}$ is locally uniformly bounded. Also, if $p_n \in \mathscr{P}(n = 1, 2, \ldots)$ and $p_n \to p$, then $p \in \mathscr{A}$ and $\operatorname{Re} p(z) \geqslant 0$ $(|z| < 1)$. Since $p(0) = p_n(0) = 1$, the minimum principle for harmonic functions [**29**, p. 271] implies that we cannot have $\operatorname{Re} p(z) = 0$, and hence $p \in \mathscr{P}$. Therefore $\mathscr{P}$ is a compact subset of $\mathscr{A}$.

The normalization $p(0) = 1$ is a convenient way of forming a compact set. Other functions in $\mathscr{A}$ having the property $\operatorname{Re} p(z) > 0$ when $|z| < 1$ are easily related to $\mathscr{P}$ through the map $p \to (p - i \operatorname{Im} p(0))/\operatorname{Re} p(0)$. The study of "positive harmonic functions" is essentially the same as the study of $\mathscr{P}$ because each such function $u$ is associated with an analytic function $p$ by $\operatorname{Re} p = u$.

The family $\mathscr{P}$ is characterized as the set of functions $p$ represented by

$$p(z) = \int_0^{2\pi} \frac{1 + e^{-it}z}{1 - e^{-it}z}\, d\mu(t) \tag{2}$$

for $|z| < 1$, where $\mu$ is a real-valued nondecreasing function on $[0, 2\pi]$ and $\mu(2\pi) - \mu(0) = 1$. Equation (2) is usually called the Herglotz formula and is due to G. Herglotz [**16**] and F. Riesz [**25**]. An easy consequence of the fact that $\operatorname{Re} P(z) > 0$ when $|z| < 1$ is that each function defined by (2) belongs to $\mathscr{P}$, given the conditions on $\mu$. The converse is more difficult to prove. An argument depending on the Helly selection theorem is given in [**9**, p. 22] and [**24**, p. 30].

Equation (2) can be rewritten in terms of probability measures on $\partial\Delta$, the boundary of $\Delta$. By a standard argument from measure theory, each real-valued nondecreasing function $\mu$ on $[0, 2\pi]$ defines a nonnegative measure $\mu^*$ on $[0, 2\pi]$. First of all, $\mu^*$ is defined on closed intervals by $\mu^*([a, b]) = \mu(b) - \mu(a)$ when $0 \leqslant a < b \leqslant 2\pi$. Then $\mu^*$ is extended to the Borel subsets of $[0, 2\pi]$ through the way such sets are related to closed intervals. If $\mu$ is normalized by $\mu(t) = \frac{1}{2}[\mu(t-) + \mu(t+)]$, then the correspondence between nondecreasing functions and nonnegative measures can be shown to be one-to-one. The condition $\mu(2\pi) - \mu(0) = 1$ is the same as $\mu^*([0, 2\pi]) = 1$; hence $\mu^*$ is a probability measure on $[0, 2\pi]$. It is convenient to regard $\mu^*$ as being defined on (suitable) subsets of $\partial\Delta$ and to write $\mu$ instead of $\mu^*$. In this way, the Herglotz formula can be stated as follows.

THEOREM. *$p \in \mathscr{P}$ if and only if*

$$p(z) = \int_{|x|=1} \frac{1 + xz}{1 - xz}\, d\mu(x), \tag{3}$$

*where $\mu$ is a probability measure on $\partial\Delta$.*

It is known that the correspondence between $\mathscr{P}$ and nondecreasing normalized functions is one-to-one, and consequently (3) gives a one-to-one map between $\mathscr{P}$ and the set of probability measures on $\partial\Delta$. Let $\Lambda$ denote the set of such measures.

Suppose that $p$, $q$, and $r$ belong to $\mathscr{P}$ and are associated with the probability measures $\mu$, $\nu$, and $\lambda$, respectively. If $p = tq + (1 - t)r$ $(0 < t < 1)$, then (3) implies that

$$\int_{|x|=1} \frac{1 + xz}{1 - xz}\, d\mu(x) = \int_{|x|=1} \frac{1 + xz}{1 - xz}\, d\eta(x),$$

where $\eta = t\nu + (1 - t)\lambda$. Since the map $\mathscr{P} \to \Lambda$ is one-to-one, $\mu = t\nu + (1 - t)\lambda$. In other words, a convex decomposition of a function in $\mathscr{P}$ corresponds to the same decomposition of the associated measure. An exercise in measure theory shows that the extreme measures of $\Lambda$ consist of the point masses, that is, the measures for which $\mu(\{x\}) = 1$ for some $x$ on $\partial\Delta$. This proves the following result.

THEOREM. *$E\mathscr{P}$ consists of the functions $p$ such that $p(z) = (1 + xz)/(1 - xz)$, where $|x| = 1$.*

The next thing to show is that the one-to-one correspondence between $\mathscr{P}$ and $\Lambda$ may be regarded as a statement about the moments of probability measures. Suppose that $p$ and $q$ belong to $\mathscr{P}$ and are associated with the probability measures $\mu$ and $\nu$, respectively,

$$p(z) = 1 + \sum_{n=1}^{\infty} p_n z^n, \text{and } q(z) = 1 + \sum_{n=1}^{\infty} q_n z^n \quad (|z| < 1).$$

Since

$$P(z) = \frac{1 + z}{1 - z} = 1 + \sum_{n=1}^{\infty} 2z^n \qquad (|z| < 1),$$

the equation (3) implies that

$$p_n = 2\int_{|x|=1} x^n\, d\mu(x) \text{ and } q_n = 2\int_{|x|=1} x^n\, d\nu(x) \quad \text{for } n = 1, 2, \ldots .$$

The functions $p$ and $q$ are equal if and only if $p_n = q_n$ $(n = 1, 2, \ldots)$. This shows that two

probability measures $\mu$ and $\nu$ are equal if and only if

$$\int_{|x|=1} x^n \, d\mu(x) = \int_{|x|=1} x^n \, d\nu(x) \qquad (n = 1,2,\ldots). \tag{4}$$

The facts about $\mathscr{P}$ just described, including the determination of $E\mathscr{P}$, are a foundation for obtaining a number of similar results for subsets of $\mathscr{A}$ in geometric function theory.

**4. Families of univalent functions.** The purpose of this section is to consider certain subsets of $\mathscr{A}$ related to conformal mapping. A function $f$ is called univalent in $D$ if it is one-to-one in $D$; that is, if $f(z_1) = f(z_2)$ and $z_1, z_2 \in D$ imply that $z_1 = z_2$. A detailed treatment of the information presented in this section and other facts about univalent functions are in the books **[9]**, **[23]**. The book **[29]** is at a more introductory level and contains several results about univalent functions, especially in Chapter 12. The classic "problem-solution" volumes **[22]** by G. Pólya and G. Szegö also contain information about univalent functions (see part IV). Both **[22]** and **[29]** are useful general references for other aspects of complex analysis, including some of the facts mentioned earlier.

DEFINITION. Let $S$ denote the subset of $\mathscr{A}$ consisting of univalent functions $f$ which satisfy $f(0) = 0$ and $f'(0) = 1$.

The so-called Koebe function $k(z) = z/(1 - z)^2$ is a member of $S$. It maps $\Delta$ onto the complement of the closed ray $\{w\colon\ w$ is real and $w \leqslant -\frac{1}{4}\}$. This follows from the mapping properties of $P(z) = (1 + z)/(1 - z)$ and the relation $k = \frac{1}{4}(P^2 - 1)$. If $f \in S$ and $0 < |x| \leqslant 1$, then $\dfrac{1}{x} f(xz)$ also is a member of $S$. In particular, $z/(1 - xz)^2$ belongs to $S$ when $|x| = 1$ and maps onto the complement of a ray. These so-called Koebe functions are extremal for a number of problems over $S$.

$S$ is not convex, since a proper convex combination of two distinct Koebe functions is not in $S$. This follows simply from the fact that such a function has two poles of order two on $\partial\Delta$. If $f$ has a pole at $\alpha(|\alpha| = 1)$ of order two, then $f(z) = g(z)/(z - \alpha)^2$, where $g$ is analytic at $\alpha$ and $g(\alpha) \neq 0$. From this relation, it can be shown that if $N$ is a neighborhood of $\alpha$, then $f$ takes on all sufficiently large complex numbers except, perhaps, those in an arbitrarily small angular region $\{w\colon\ |\arg w - \alpha| < \varepsilon\}$, for $z \in N \cap \Delta$. It follows that a function with two poles of order two on $\partial\Delta$ takes on (most) large complex numbers at least twice, as $z$ varies in $\Delta$.

Univalence is readily lost through convex combinations of functions. One way to construct suitable examples is to rely on the result that a univalent function has a non-vanishing derivative [**29**, p. 299]. (More generally, non-constant analytic functions have the property that they are $n$ to 1 locally, if the derivative has a zero of order $n - 1$ at a point [**2**, p. 133].) For example, without the normalization $f'(0) = 1$, the average of two univalent functions may be constant, as $\frac{1}{2}z + \frac{1}{2}(-z) = 0$ illustrates. Even with the normalizations of $S$, this approach provides suitable examples. For example, let $f$ be a member of $S$ with $\operatorname{Re} f'(z_0) = 0$ and $|z_0| < 1$. Define $g$ by $g(z) = \overline{x f(x\bar{z})}$, where $x = \exp(2i \arg z_0)$. Then $g \in S$, $g'(z_0) = \overline{f'(z_0)}$ and $f'(z_0) + g'(z_0) = 0$. Thus, $h = \frac{1}{2}f + \frac{1}{2}g$ has a vanishing derivative at $z_0$, has the normalizations of $S$ but is not univalent. (It is not difficult to construct functions in $S$ such that $\operatorname{Re} f'(z_0) = 0$ for some $z_0$.) An even more striking example, due to A. W. Goodman [**11**], shows that there are two functions $f$ and $g$ in $S$ such that $\frac{1}{2}f + \frac{1}{2}g$ takes on some value an infinite number of times in $\Delta$.

It is well known that $S$ is locally uniformly bounded and, in fact, $|f(z)| \leqslant |z|/(1 - |z|)^2$ when $f \in S$ and $|z| < 1$ [**29**, p. 326]. A theorem of A. Hurwitz asserts that if $\{f_n\}$ is a sequence of functions which are analytic and univalent in a region $D$ and if $f_n \to f$, then either $f$ is univalent in $D$ or $f$ is constant [**29**, p. 311]. If $f_n \in S$ and $f_n \to f$, then $f'(0) = 1$, and thus $f$ must be univalent in $\Delta$. This verifies that $S$ is compact.

Consider next subsets of $S$ consisting of starlike and convex mappings. A set $D$ of complex numbers is called starlike with respect to $w_0$ provided that $tw + (1 - t)w_0 \in D$ when $w \in D$ and

$0 < t < 1$. In other words, if a point is in $D$, then so is the line segment connecting that point to $w_0$.

Definition. Let $S^*$ denote the subset of $S$ consisting of functions $f$ for which $f(\Delta)$ is starlike with respect to 0.

The Koebe functions belong to $S^*$ since they map $\Delta$ onto the complement of a ray. Because $S^* \subset S$, the set $S^*$ is locally uniformly bounded. Also, if $f_n \in S^*$ and $f_n \to f$, then $f \in S^*$. One way to show this is to appeal to the geometric fact that the limit of the starlike sets $f_n(\Delta)$ is a starlike set. An alternative argument is based on the condition (5) given below. Therefore $S^*$ is compact. The example of a proper convex combination of distinct Koebe functions also serves to show that $S^*$ is not convex. The property of belonging to $S^*$ is easily lost through convex combinations; thus it is somewhat unexpected that the techniques involving convexity considerations and described later are so effective with $S^*$.

The family $S^*$ is analytically characterized by the following conditions: $f \in \mathscr{A}$, $f(0) = 0$, $f'(0) = 1$ and

$$\operatorname{Re}\left\{\frac{zf'(z)}{f(z)}\right\} > 0 \quad (|z| < 1) \tag{5}$$

[29, p. 332–333]. The inequality (5) expresses the local fact that the angle of the vector $f(re^{i\theta})$ increases with $\theta$. This follows from the relations

$$\begin{aligned}\frac{\partial}{\partial\theta}\arg f(re^{i\theta}) &= \frac{\partial}{\partial\theta}\operatorname{Im}\log f(re^{i\theta})\\ &= \operatorname{Im}\left\{\frac{\partial}{\partial\theta}\log f(re^{i\theta})\right\} = \operatorname{Im}\left\{\frac{f'(re^{i\theta})}{f(re^{i\theta})}ire^{i\theta}\right\} = \operatorname{Re}\left\{\frac{re^{i\theta}f'(re^{i\theta})}{f(re^{i\theta})}\right\}.\end{aligned}$$

The argument for the characterization uses the fact that if $f \in S^*$ and $0 < r < 1$, then $f[\{z: |z| < r\}]$ is starlike with respect to 0. The last relation is a consequence of an appropriate application of Schwarz's lemma. The argument also depends on showing that (5) together with $f(0) = 0$ and $f'(0) \neq 0$ implies that the curve $w = f(re^{i\theta})$, $0 \leqslant \theta \leqslant 2\pi$, winds around the origin exactly once with an increasing angle. This property of the curve implies that $f$ is univalent in $\{z: |z| < r\}$ for each $r$ with $0 < r < 1$.

Condition (5) implies a close relation between $S^*$ and $\mathscr{P}$. If

$$p(z) = \frac{zf'(z)}{f(z)} \quad (|z| < 1), \tag{6}$$

it follows that $p \in \mathscr{P}$ if and only if $f \in S^*$. Moreover, (6) defines a one-to-one correspondence between $\mathscr{P}$ and $S^*$. The argument for the latter fact requires verifying that (6) may be uniquely solved for $f$. To do this, first note that if $f \in S^*$, then $f(z)/z$ is analytic and non-vanishing in $\Delta$. Thus $\log\dfrac{f(z)}{z}$ is analytic in $\Delta$, and

$$\frac{d}{dz}\log\frac{f(z)}{z} = \frac{p(z)-1}{z}.$$

This implies that

$$f(z) = z\exp\left\{\int_0^z \frac{p(w)-1}{w}\,dw\right\}. \tag{7}$$

A use of (3) in (7), followed by an interchange of the order of integrations, yields the representation

$$f(z) = z\exp\left\{-2\int_{|x|=1}\log(1-xz)\,d\mu(x)\right\} \quad (|z| < 1) \tag{8}$$

for function in $S^*$, where $\mu \in \Lambda$.

DEFINITION. Let $K$ denote the subset of $S$ consisting of the functions $f$ for which $f(\Delta)$ is convex.

The function $f(z) = z/(1 - z)$ belongs to $K$, and it maps $\Delta$ one-to-one onto the half plane $\{w: \operatorname{Re} w > -\frac{1}{2}\}$. Since $K \subset S$, it follows that $K$ is locally uniformly bounded, and a geometric argument can be given to verify that $K$ is closed. Therefore $K$ is compact. Also, $K$ is not convex. To see this, note that $z/(1 - xz)$ belongs to $K$ whenever $|x| \leqslant 1$, and $\sum_{k=1}^{3} t_k z/(1 - x_k z)$ is not univalent in $\Delta$ when $x_k$ are distinct, $|x_k| = 1$, $t_k > 0$ and $\sum_{k=1}^{3} t_k = 1$. The non-univalence follows from the local behavior of a function at a simple pole. The function described by the sum takes on certain large values at least twice in $\Delta$.

The family $K$ also has an analytic characterization. Namely, $f \in K$ if and only if $f \in \mathscr{A}$, $f(0) = 0$, $f'(0) = 1$ and

$$\operatorname{Re}\left\{\frac{zf''(z)}{f'(z)} + 1\right\} > 0 \qquad (|z| < 1) \tag{9}$$

[29, p. 336]. As a local condition, (9) is equivalent to the assertion that the tangent to the curve $w = f(re^{i\theta}), 0 \leqslant \theta \leqslant 2\pi$, turns counterclockwise as $\theta$ increases. This follows because the angle of the tangent to this curve is $\arg\{izf'(z)\}$ and

$$\frac{\partial}{\partial\theta} \arg\{ire^{i\theta}f'(re^{i\theta})\} = \frac{\partial}{\partial\theta} \operatorname{Im}\log\left[ire^{i\theta}f'(re^{i\theta})\right] = \operatorname{Re}\left\{\frac{re^{i\theta}f''(re^{i\theta})}{f'(re^{i\theta})} + 1\right\}.$$

A comparison of (5) and (9) shows that the map $f \to g$ defined by $g(z) = \int_0^z \frac{f(w)}{w}\,dw$ is a one-to-one correspondence from $S^*$ onto $K$. Note that the normalizations for $S^*$ and $K$ imply the equivalent relation $f(z) = zg'(z)$.

**5. The Krein-Milman theorem.** The convex hull of a set $\mathscr{F}$ is defined as the intersection of all convex sets containing $\mathscr{F}$. Likewise, the closed convex hull of $\mathscr{F}$, which is denoted by $H\mathscr{F}$, is defined as the intersection of all closed convex sets containing $\mathscr{F}$. If $\mathscr{F} \subset \mathscr{A}$, then $f \in H\mathscr{F}$ provided that $f$ is the limit (in the topology of $\mathscr{A}$) of functions having the form

$$\sum_{k=1}^{n} t_k f_k, \tag{10}$$

where $f_k \in \mathscr{F}$, $t_k \geqslant 0$ and $\sum_{k=1}^{n} t_k = 1$. This is a consequence of the fact that the functions defined by (10), with varying $t_k$, $f_k$ and $n$, form the smallest convex set containing $\mathscr{F}$. The set $H\mathscr{F}$ is compact when $\mathscr{F}$ is compact. To prove this, it is only necessary to show that $H\mathscr{F}$ is locally uniformly bounded. This follows from the same property of $\mathscr{F}$; that is, if $|f(z)| \leqslant M$ when $|z| \leqslant r$ and $f \in \mathscr{F}$, the same inequality holds when $f \in H\mathscr{F}$ by first applying the triangle inequality to (10).

The Krein-Milman theorem is of fundamental importance in questions concerning extreme points and linear extremal problems. It holds in any locally convex linear topological space $X$ and consists of the following assertions [7, p. 439–446].

THEOREM. *Suppose that $Y$ is a compact subset of $X$.*

(11.1) *If $Y$ is non-empty, then $EY$ is non-empty.*

(11.2) $HEY = HY$.

(11.3) *If $HY$ is compact, then $EHY \subset Y$.*

As remarked earlier, the compactness of a subset $\mathscr{F}$ of $\mathscr{A}$ implies the compactness of $H\mathscr{F}$. This fact and (11.3) imply that $EH\mathscr{F} \subset \mathscr{F}$ when $\mathscr{F}$ is a compact subset of $\mathscr{A}$.

In the following way, the Herglotz formula yields (11.3) in the case of the family $\mathscr{P}$. If $p(z)$

$= \int_{|x|=1} \frac{1+xz}{1-xz}\, d\mu(x)$ and $\mu \in \Lambda$, then $p$ is a limit of functions of the form $\sum_{k=1}^{n} t_k \frac{1+x_k z}{1-x_k z}$, where $|x_k| = 1$, $t_k \geqslant 0$ and $\sum_{k=1}^{n} t_k = 1$. This follows from the approximation obtained by replacing $\mu$ by a convex average of point masses at $x_k$. This is the same as the assertion that the set of all such finite sums is a dense subset of $\mathscr{P}$. (One proof is based on the weak* compactness of $\Lambda$.)

Suppose that $X$ and $Y$ are two linear topological spaces. A mapping $\mathscr{L}$: $X \to Y$ is called a linear homeomorphism if it is a homeomorphism of $X$ onto $Y$ and satisfies $\mathscr{L}(\alpha f + \beta g) = \alpha \mathscr{L}(f) + \beta \mathscr{L}(g)$ for all scalars $\alpha$ and $\beta$ and all vectors $f$ and $g$. If $\mathscr{A}_0$ denotes the subspace of $\mathscr{A}$ consisting of functions $f$ such that $f(0) = 0$, then $g(z) = \int_0^z \frac{f(w)}{w}\, dw$ defines a linear homeomorphism of $\mathscr{A}_0$ onto $\mathscr{A}_0$. Let $\mathscr{L}$ denote this mapping. To show the continuity of $\mathscr{L}$, verify first that if $f_n \to f$ and $g_n(z) = \int_0^z \frac{f_n(w)}{w}\, dw$, then $g_n \to g$. Also, $f(z) = zg'(z)$, and Weierstrass' result implies that if $g_n \to g$, then $f_n \to f$. Earlier, it was shown that $\mathscr{L}(S^*) = K$, and the general properties of linear homeomorphisms yield $\mathscr{L}(HS^*) = HK$ and $\mathscr{L}(EHS^*) = EHK$.

**6. Extreme points of HS* and HK.** In this section, $HS^*$, $HK$ and their set of extreme points are determined. First a family closely related to $S^*$ is studied.

Let $\mathscr{H}$ denote the subset of $\mathscr{A}$ consisting of functions $f$ for which $f(0) = 1$ and $f(\Delta) \subset F(\Delta)$, where $F(z) = 1/(1-z)^2$. (We say that $f$ is subordinate to $F$.) The set $\mathscr{H}$ is compact. To see that $\mathscr{H}$ is closed, note that if $f_n \in \mathscr{H}$ and $f_n \to f$, then $f(0) = 1$ and $f(\Delta) \subset \overline{F(\Delta)}$. The last two relations and the fact that non-constant analytic functions are open mappings [**2**, p. 132] show that $f(\Delta) \subset F(\Delta)$. Also, $\mathscr{F}$ is locally uniformly bounded because of Schwarz's lemma applied to the analytic function $F^{-1} \circ f$. This yields $|f(z)| \leqslant 1/(1-|z|)^2$ when $|z| < 1$ and $f \in \mathscr{H}$.

If $\mathscr{G}$ denotes the subset of $\mathscr{A}$ described by the conditions $g(0) = 1$ and $\operatorname{Re} g(z) > \frac{1}{2}(|z| < 1)$, then $f \in \mathscr{H}$ if and only if $f = g^2$ where $g \in \mathscr{G}$. Also, $g \in \mathscr{G}$ if and only if $2g - 1 \in \mathscr{P}$. The result about $E\mathscr{P}$ implies that $E\mathscr{G}$ consists of the functions $g(z) = 1/(1 - xz)$ where $|x| = 1$.

Suppose that $f \in EH\mathscr{H}$. Then, by (11.3), $f \in \mathscr{H}$ and hence $f = g^2$, where $g \in \mathscr{G}$. To show that $g \in E\mathscr{G}$, assume that $g \notin E\mathscr{G}$. Then

$$g = tg_1 + (1-t)g_2, \quad \text{where} \quad 0 < t < 1, \qquad g_1, g_2 \in \mathscr{G} \text{ and } g_1 \neq g_2.$$

This implies that $f = tf_1 + (1-t)f_2$, where $f_1 = gg_1$ and $f_2 = gg_2$. This contradicts $f \in EH\mathscr{H}$ since $gh \in \mathscr{H}$ whenever $g, h \in \mathscr{G}$. This latter fact may be regarded as a statement about complex numbers; namely, if $\operatorname{Re} w_1 > \frac{1}{2}$ and $\operatorname{Re} w_2 > \frac{1}{2}$, then $\operatorname{Re}\sqrt{w_1 w_2} > \frac{1}{2}$ (this follows by expressing $w_1$ and $w_2$ in polar form and later using the concavity of $u(\theta) = \log\cos\theta$). Therefore, $EH\mathscr{H}$ consists only of functions having the form $f(z) = 1/(1-xz)^2$, where $|x| = 1$. It follows from (11.2) that each function $f$ in $H\mathscr{H}$ is represented by

$$f(z) = \int_{|x|=1} \frac{1}{(1-xz)^2}\, d\mu(x) \quad (|z| < 1), \tag{12}$$

where $\mu \in \Lambda$.

Suppose that $f \in S^*$. According to (8), $f(z) = z\exp\{2q(z)\}$, where

$$q(z) = \int_{|x|=1} \log\frac{1}{1-xz}\, d\mu(x) \tag{13}$$

and $\mu \in \Lambda$. The function $Q(z) = \log\frac{1}{1-z}$ belongs to $K$, as can be shown by a computation which verifies (9). Since (13) expresses the complex number $q(z)$ as a convex average of values of $Q$, this implies that $q(\Delta) \subset Q(\Delta)$. Therefore the function $f(z)/z$ belongs to $\mathscr{H}$. An application

of (12) shows that

$$f(z) = \int_{|x|=1} \frac{z}{(1-xz)^2}\, d\mu(x) \tag{14}$$

for some $\mu \in \Lambda$.

Let $\mathscr{F}$ denote the set of functions represented by (14), where $\mu \in \Lambda$. It follows that $\mathscr{F}$ is convex, and it can be shown (using the weak* compactness of $\Lambda$) that $\mathscr{F}$ is closed. Therefore $HS^* \subset \mathscr{F}$. Conversely, $\mathscr{F} \subset HS^*$ since the Koebe functions $z/(1-xz)^2(|x|=1)$ belong to $S^*$ and $\mathscr{F}$ is the closed convex hull of these functions. Therefore, (14) characterizes the functions $f$ in $HS^*$.

The map $HS^* \to \Lambda$ given by (14) is one-to-one. This is a consequence of the fact that the coefficients of the power series for such functions are given by $a_n = n\int_{|x|=1} x^{n-1}\, d\mu(x)$ $(n = 1,2,\dots)$ and the result about the moments of probability measures described in Section 3. This one-to-one property shows that $EHS^*$ is the set of Koebe functions, with the same argument given for $E\mathscr{P}$.

THEOREM. *$HS^*$ consists of functions $f$ given by* (14), *where $\mu$ varies over the probability measures on $\partial\Delta$. $EHS^*$ consists of the Koebe functions $f(z) = z/(1-xz)^2$, where $|x| = 1$.*

The results about $HS^*$ transform to equivalent facts about $HK$ through the linear homeomorphism between $S^*$ and $K$, as described in Section 5. This leads to the following result.

THEOREM. *$HK$ consists of the functions $f$ represented by*

$$f(z) = \int_{|x|=1} \frac{z}{1-xz}\, d\mu(x) \quad (|z|<1), \tag{15}$$

*where $\mu \in \Lambda$, and $EHK^*$ consists of the functions $f(z) = z/1 - xz$, where $|x| = 1$.*

The kind of argument given above to obtain $EHS^*$ and $EHK$ has been successfully applied to several families in geometric function theory, including the so-called typically real functions, the close-to-convex functions, and the functions with bounded boundary rotation [4], [5]. (In the case of functions of boundary rotation at most $k\pi$, the problem remains open when $2 < k < 4$.) The argument given about $H\mathscr{K}$ is a particular case of a more general argument which shows that the set of extreme points of the closed convex hull of the functions subordinate to $F(z) = \left(\dfrac{1+cz}{1-z}\right)^\alpha$ consists of the functions $F(xz)$ $(|x|=1)$, whenever $|c| \leqslant 1$, $c \neq -1$ and $\alpha \geqslant 1$ [4]. The results about $EHS^*$ and $EHK$ were first proved by L. Brickman, D. R. Wilken, and the author in [5].

**7. Applications to extremal problems.** The question of finding the extreme points of a given set is interesting in itself, but this question is especially important in complex analysis because of its relation to the solution of extremal problems. In this section, a variety of extremal problems are described which can be solved using the information about the extreme points of families obtained earlier.

DEFINITIONS. The term functional means a complex-valued function defined on $\mathscr{A}$. A functional $L$ is called continuous (on $\mathscr{A}$) if $L(f_n) \to L(f)$ when $f_n \in \mathscr{A}$ and $f_n \to f$. A functional $L$ is called linear if

$$L(\alpha f + \beta g) = \alpha L(f) + \beta L(g)$$

when $\alpha, \beta \in \mathbb{C}$ and $f, g \in \mathscr{A}$.

If $|z_0| < 1$ and $n$ is a nonnegative integer, then $L(f) = f^{(n)}(z_0)$ is a continuous linear functional. The continuity of this functional follows from a repeated application of Weierstrass' theorem, which shows that if $f_k \to f$, then $f_k^{(n)} \to f^{(n)}$ for $n = 1,2,\dots$. Another example of a continuous linear functional is given by the $n$th coefficient of the power series of functions in $\mathscr{A}$.

More generally, if $b_0, b_1, \dots, b_n$ is a finite sequence of complex numbers and $L(f) = \sum_{k=0}^{n} b_k a_k$ when $f(z) = \sum_{k=0}^{\infty} a_k z^k$ ($|z| < 1$), then $L$ is a continuous linear functional. This may be extended to infinite sequences $\{b_k\}$ provided that $\overline{\lim}_{k\to\infty} \sqrt[k]{|b_k|} < 1$, and then $L(f) = \sum_{k=0}^{\infty} b_k a_k$ characterizes the form of each continuous linear functional [**9**, p. 280].

A fundamental area of research in classical complex analysis is the solution of extremal problems. For the family $\mathscr{A}$, one general form for extremal problems is described as follows: Find $\sup\{\operatorname{Re} J(f) : f \in \mathscr{F}\}$ where $J$ is a functional and $\mathscr{F} \subset \mathscr{A}$. If $\mathscr{F}$ is compact and $J$ is continuous, then the range of $\operatorname{Re} J$ is a compact set of real numbers, and thus $\operatorname{Re} J(f)$ has a maximum over $\mathscr{F}$. Any function $f$ in $\mathscr{F}$ for which $\operatorname{Re} J(f) = \max\{\operatorname{Re} J(g) : g \in \mathscr{F}\}$ is called an extremal function or a solution to the extremal problem.

When the continuous functional is also linear, more can be said about extremal functions, and this relates to extreme points.

THEOREM. *If $\mathscr{F}$ is a compact subset of $\mathscr{A}$ and $L$ is a continuous linear functional on $\mathscr{A}$, then*

$$\max_{f \in EH\mathscr{F}} \operatorname{Re} L(f) = \max_{f \in \mathscr{F}} \operatorname{Re} L(f) = \max_{f \in H\mathscr{F}} \operatorname{Re} L(f). \tag{16}$$

The last two maxima in (16) exist because $\mathscr{F}$ and $H\mathscr{F}$ are compact. However, since $EH\mathscr{F}$ is not always compact, the existence of the first maximum is not immediately evident. Because of the definition of $H\mathscr{F}$ and (11.3), $EH\mathscr{F} \subset \mathscr{F} \subset H\mathscr{F}$. Therefore, if we are interested in the problem of maximizing $\operatorname{Re} L$ over $\mathscr{F}$, (16) permits us to solve the problem only over the smaller set $EH\mathscr{F}$. This becomes a useful reduction when $EH\mathscr{F}$ can be found and is considerably smaller than $\mathscr{F}$. At the same time, (16) shows that the extremal problem over $\mathscr{F}$ has the same maximum over the larger set $H\mathscr{F}$.

Here is an outline of the proof of (16). It is a simple abstract argument based on the Krein-Milman theorem. If

$$M = \max\{\operatorname{Re} L(f) : f \in H\mathscr{F}\} \quad \text{and} \quad \mathscr{G} = \{f \in H\mathscr{F} : \operatorname{Re} L(f) = M\},$$

it can be shown that $\mathscr{G}$ is compact and convex. By (11.1) and (11.3), $\mathscr{G}$ has an extreme point, say $g$. It follows that $g$ is an extreme point of $H\mathscr{F}$, because the relations

$$g = th + (1-t)k, \qquad 0 < t < 1 \text{ and } h, k \in H\mathscr{F}$$

imply that

$$M = \operatorname{Re} L(g) = t \operatorname{Re} J(h) + (1-t) \operatorname{Re} J(k) \leqslant tM + (1-t)M = M,$$

and thus $h, k \in \mathscr{G}$. In other words, there is an element of $EH\mathscr{F}$ in the solution set $\mathscr{G}$.

Here are two applications of (16), given to $\mathscr{P}$ and to $S^*$. First, recall that $E\mathscr{P}$ consists of the functions $P(xz)$, where $|x| = 1$ and $P(z) = 1 + \sum_{n=1}^{\infty} 2z^n$. Therefore, the problem of finding $\max\{\operatorname{Re} a_n : p \in \mathscr{P}\}$, where $p(z) = 1 + \sum_{n=1}^{\infty} a_n z^n$, is reduced to finding $\max_{|x|=1} \operatorname{Re}(2x^n)$, which equals 2. Second, if $f \in S^*$ and $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$, then $\operatorname{Re} a_n \leqslant n$, because

$$\max\{\operatorname{Re} a_n : f \in EHS^*\} = \max\{\operatorname{Re}(nx^{n-1}) : |x| = 1\}.$$

In general, linear extremal problems over $\mathscr{P}$, $S^*$, or $K$ are reduced to specific computational problems of maximizing the real part of certain functions of $x$ where $|x| = 1$. Several other families have the property that their set of extreme points may be parametrized by one or more real or complex variables over specific sets, and thus linear problems over these families also are reduced to specific computations.

The theorems determining $HS^*$ and $HK$ and the Herglotz formula are examples of Choquet's theorem [**21**], which gives the representation of a continuous linear functional in terms of integration with respect to a probability measure on the extreme points of a set, for fairly general locally convex spaces. Since the extreme points of $HS^*$, $HK$ and $\mathscr{P}$ are parametrized over $\partial\Delta$, the representation of the functional $L(f) = f(z)$ ($|z| < 1$) takes on the forms obtained earlier. Other examples of families in geometric function theory have their extreme points parametrized

on sets such as $[0,1]$ or $\partial\Delta \times \partial\Delta$, and thus these families illustrate Choquet's theorem with representations involving integrals over such sets.

A real-valued function $J$ defined on a vector space $X$ is called convex if

$$J(tx + (1-t)y) \leqslant tJ(x) + (1-t)J(y)$$

when $x, y \in X$ and $0 < t < 1$. The argument used to prove (16) shows that

$$\max\{J(f) : f \in EH\mathscr{F}\} = \max\{J(f) : f \in \mathscr{F}\} = \max\{J(f) : f \in H\mathscr{F}\},$$

when $J$ is continuous and convex on $\mathscr{A}$ and $\mathscr{F}$ is a compact subset of $\mathscr{A}$. For example, this applies to $J(f) = |a_n|$, where $a_n$ is the $n$th coefficient of the power series for $f$, or more generally to $J = |L|$ where $L$ is a continuous linear functional.

Another interesting convex functional is given by

$$J(f) = \left\{\int_0^{2\pi} |f^{(n)}(re^{i\theta})|^\lambda \, d\theta\right\}^{1/\lambda},$$

where $0 < r < 1$, $n$ is a nonnegative integer, and $\lambda \geqslant 1$. The convexity of $J$ follows from the Minkowski inequality. One application of these relations is that if $f \in K$ and $F(z) = z/(1-z)$, then

$$\int_0^{2\pi} |f^{(n)}(re^{i\theta})|^\lambda \, d\theta \leqslant \int_0^{2\pi} |F^{(n)}(re^{i\theta})|^\lambda \, d\theta \tag{17}$$

for $0 < r < 1$, $\lambda \geqslant 1$ and $n = 0, 1, \ldots$. The argument depends on the fact that $EHK$ consists of the functions $F(xz)/x$ $(|x| = 1)$, and hence $J$ is constant on $EHK$, as the change of variables $xe^{i\theta} \to e^{i\theta}$ shows. This result is contained in [**17**].

Another area of application concerns linear operators of order zero. In the context of the space $\mathscr{A}$, $\mathscr{L}$ is a continuous linear operator provided that $\mathscr{L}$ is a mapping of $\mathscr{A}$ into $\mathscr{A}$ such that $\mathscr{L}(\alpha f + \beta g) = \alpha\mathscr{L}(f) + \beta\mathscr{L}(g)$ when $\alpha, \beta \in \mathbb{C}$ and $f, g \in \mathscr{A}$, and $\mathscr{L}(f_n) \to \mathscr{L}(f)$ when $f_n \in \mathscr{A}$ and $f_n \to f$. We say that $\mathscr{L}$ is of order zero if, in addition,

$$\mathscr{L}[f(\cdot)]|_{xz} = \mathscr{L}[f(\cdot)|_{xz}]$$

when $f \in \mathscr{A}$ and $0 < |x| \leqslant 1$; that is, $\mathscr{L}$ commutes with the change of variables $z \to xz$. Some examples of continuous linear operators of order zero are those which assign to $f$ the functions given by:

$$zf'(z), [zf(z)]', \int_0^z \frac{f(w) - f(0)}{w} \, dw, \frac{1}{z}\int_0^z f(w) \, dw,$$

and the $n$th partial sum of the power series for $f$. It is easy to verify that these operators have the required properties. For example, the continuity of the first two operators is a consequence of Weierstrass' theorem. In general, such operators are characterized by $\mathscr{L}[f(z)] = \sum_{n=0}^{\infty} a_n b_n z^n$, where $f(z) = \sum_{n=0}^{\infty} a_n z^n$ and the sequence $\{b_n\}$ satisfies $\overline{\lim}_{n\to\infty} |b_n|^{1/n} \leqslant 1$ (see [**18**]).

Suppose that $F \in \mathscr{A}$, $\mathscr{L}$ is a continuous linear operator of order zero, and $\mathscr{F}$ is a compact subset of $\mathscr{A}$ for which $EH\mathscr{F}$ consists of the functions of the form $F(xz)$ where $|x| = 1$. If $f \in \mathscr{F}$, then the Krein-Milman theorem implies that $f(z) = \int_{|x|=1} F(xz)\, d\mu(x)$, where $\mu \in \Lambda$. This integral can be approximated by suitable finite sums $\sum t_k F(x_k z)$. The linearity of $\mathscr{L}$ and the continuity of $\mathscr{L}$ yield the formula

$$\mathscr{L}[f](z) = \int_{|x|=1} \mathscr{L}[F(\cdot)|_{xz}] \, d\mu(x).$$

Since $\mathscr{L}$ is of order zero, this formula can be written

$$\mathscr{L}[f](z) = \int_{|x|=1} \mathscr{L}[F(\cdot)]|_{xz} \, d\mu(x). \tag{18}$$

Equation (18) expresses the complex number $\mathscr{L}[f](z)$ as a convex average of the numbers $\mathscr{L}[F(\cdot)]|_{xz}$. Therefore, $\mathscr{L}[f](z)$ belongs to the closed convex hull of the set of numbers $\mathscr{L}[F](w)$ where $|w| \leqslant |z|$. (We say that $\mathscr{L}[f]$ is hull subordinate to $\mathscr{L}[F]$.)

As a specific example, (18) applies where $\mathscr{F} = \mathscr{P}$ and $\mathscr{L}$ is the $n$th partial sum of a power series. Let $p_n$ be the $n$th partial sum of the power series of a function $p$ in $\mathscr{P}$ and let $P_n$ be the $n$th partial sum of $P$. If $0 < r < 1$, then $p_n[\{z : |z| \leqslant r\}]$ is contained in the closed convex hull of $P_n[\{z : |z| \leqslant r\}]$. In particular, the largest value of $r_n$ such that $\operatorname{Re} p_n(z) \geqslant 0$ when $|z| \leqslant r_n$ can be obtained by a specific computation using $P_n$. Note that $P(z) = (1 + z)/(1 - z)$, and this implies that $P_n(z) = (1 + z - 2z^{n+1})/(1 - z)$ for $z \neq 1$. This yields the classical result that $\operatorname{Re} p_n(z) \geqslant 0$ when $|z| \leqslant 1/2$ and $n = 1, 2, \ldots$ .

The diversity of continuous linear operators of order zero and the interesting examples of compact families $\mathscr{F}$ such that $EH\mathscr{F}$ consists of the functions of the form $F(xz)$, where $|x| = 1$, provide numerous applications of the method just described. The idea of linear operators of order zero in relation to $\mathscr{A}$ was introduced by R. M. Robinson in [**26**], and the connection of this idea to extreme points was presented by the author in [**18**].

Extremal problems, some of which are quite nonlinear, can be solved using ideas about extreme points in a less direct way. Here are two such examples.

Suppose that $J$ and $L$ are continuous linear functionals and $L(p) \neq 0$ when $p \in \mathscr{P}$. Also, let $\mathscr{P}_0$ denote the set of functions of the form

$$p(z) = t\frac{1 + xz}{1 - xz} + (1 - t)\frac{1 + yz}{1 - yz}, \quad \text{where} \quad 0 \leqslant t \leqslant 1 \quad \text{and} \quad |x| = |y| = 1.$$

Then

$$\max_{p \in \mathscr{P}} \operatorname{Re}\left\{\frac{J(p)}{L(p)}\right\} = \max_{p \in \mathscr{P}_0} \operatorname{Re}\left\{\frac{J(p)}{L(p)}\right\}. \tag{19}$$

This result is a particular case of a more general theorem due to St. Ruscheweyh [**27**]. The relation (19) provides a reduction of such extremal problems to a computational problem involving the parameters $t$, $x$ and $y$. The proof of (19) depends on showing that $J/L$ has the same range over $\mathscr{P}_0$ as it has over $\mathscr{P}$. To show this, suppose that $p \in \mathscr{P}$ and $\lambda = J(p)/L(p)$. Then, the continuous linear functional $N = J - \lambda L$ equals 0 at $p$. Since each extremal problem of the form $\max\{\operatorname{Re} e^{i\theta}N(p) : p \in \mathscr{P}\}$ ($\theta$ real) has a solution in $E\mathscr{P}$, 0 belongs to the closed convex hull of the set of values that $N$ achieves at the extreme points of $\mathscr{P}$. The set $\{N(P(xz)) : |x| = 1\}$ can be shown to be an analytic curve, and, in particular, it is connected. It follows from the two-dimensional case of a theorem of Carathéodory [**10**, p. 35] that the number 0 is the convex average of at most two elements of that set. This implies that $\lambda \in \frac{J}{L}(\mathscr{P}_0)$.

The next example concerns the fact that the form of solutions to the extremal problems $\max_{f \in \mathscr{F}} \operatorname{Re} \Phi[f(\zeta), f'(\zeta), \ldots, f^{(n)}(\zeta)]$, with $\zeta$ fixed and $|\zeta| < 1$, can be identified for fairly general functions $\Phi$. As in previous examples, a reduction takes place, and it is an effective step in solving such problems at least when $n$ is small. The method is applicable if either $\mathscr{F}$ is convex (and compact) with a simple set of extreme points, or if $\mathscr{F}$ is related to such a family through an analytic mapping. The idea is first to apply extreme-point techniques to a convex family in order to identify functions corresponding to boundary points of the family's region of variability, the set of points $(f(\zeta), f'(\zeta), \ldots, f^{(n)}(\zeta))$ with varying $f$. Then that information is transferred to equivalent facts about the boundary points of the region of variability of the related family. This is possible since the two regions often are related by a homeomorphism. For example, this argument is directly applicable to $\mathscr{P}$ and to the family described by (13). The second family, in turn, is related to $S^*$ by $f(z) = z \exp[2q(z)]$, and this map generates the appropriate homeomorphism. The details of these arguments and comments about related results, including coefficient

problems, are given in [**19**]. When $\mathscr{F}$ is not convex, the solution to such general extremal problems usually gives a number which is larger over $H\mathscr{F}$ than over $\mathscr{F}$.

For families having a large set of extreme points, it is less likely that the methods outlined here will be successful. In any case, the general theory is applicable. For example, to solve the problem $\max_{f \in S}|a_n|$, where $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$, it suffices to consider functions in $EHS$. A few facts about $EHS$ are described in [**9**, pp. 286ff.], including the result that if $f \in ES$, then $f(\Delta)$ is the complement of a continuous curve $w = w(t)$, $0 \leqslant t < 1$, whose modulus strictly increases and tends to $\infty$ as $t \to 1$. Such results, naturally, are of interest with regard to the famous Bieberbach conjecture, which states that if $f \in S$ and $f(z) = z + \sum_{n=2}^{\infty} a_n z^n$, then $|a_n| \leqslant n (n = 2, 3, \ldots)$. Note that, for the Koebe functions, $|a_n| = n$. (This conjecture has recently been proved by L. de Branges.)

**8. Families having ranges in a convex region.** For the families discussed next the determination of the extreme points is a much more difficult problem, requiring further advanced ideas, including results from the theory of $H^p$ spaces [**8**]. The presentation in this section is generally more descriptive than those given earlier.

Let $D$ be a convex region and let $w_0 \in D$. Let $\mathscr{F}$ denote the subset of $\mathscr{A}$ for which $f(\Delta) \subset D$ and $f(0) = w_0$. Then $\mathscr{F}$ is a compact, convex subset of $\mathscr{A}$. The compactness follows primarily from Schwarz's lemma, applied to $F^{-1} \circ f$, where $f \in \mathscr{F}$ and $F$ is a one-to-one map of $\Delta$ onto $D$ with $F(0) = w_0$. The existence of $F$ depends on the Riemann mapping theorem [**2**, p. 230]. A description of $E\mathscr{F}$ shall be given for certain specific regions $D$ and for general regions with sufficiently smooth boundaries.

The set $H^p$ is defined as the subset of $\mathscr{A}$ consisting of functions $f$ such that the integrals $\int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta$ are bounded on the interval $\{r : 0 < r < 1\}$. Generally it is assumed that $D \neq \mathbb{C}$ and $D$ is not a half plane. A consideration of the support lines to $D$ implies that $f(\Delta)$ is contained in an angular wedge of opening less than $\pi$ or in an open strip. From this geometric restriction on the range of $f$, it can be shown that $f \in H^1$ (see [**1**]). This permits the application of a number of facts about $H^1$. In particular, $f$ may be represented by the Poisson formula

$$f(re^{i\theta}) = \frac{1}{2\pi} \int_0^{2\pi} P(r, \theta - t) f(e^{it}) \, dt, \tag{20}$$

where

$$P(r, \theta) = \frac{1 - r^2}{1 - 2r\cos\theta + r^2} \tag{21}$$

and

$$f(e^{i\theta}) = \lim_{r \to 1} f(re^{i\theta}). \tag{22}$$

The limit in (22) exists for almost all $\theta$ on $[0, 2\pi)$ and $f(e^{i\theta})$ is Lebesgue integrable.

The results about $E\mathscr{F}$ are given in terms of the "boundary function" $f(e^{i\theta})$. A critical step in such arguments is the construction of a function in terms of its boundary values through (20). Formula (20) expresses $f(z)$ as a convex average of the numbers $f(e^{it})$, since $P(r, \theta) \geqslant 0$ and $\frac{1}{2\pi} \int_0^{2\pi} P(r, \theta) \, d\theta = 1$. This permits the conclusion that $f \in \mathscr{F}$ essentially from $f(e^{i\theta}) \in \overline{D}$ for almost all $\theta$. The actual construction of suitable boundary functions often depends on the use of harmonic measure [**31**, pp. 111ff.].

One general result asserts that if $f \in \mathscr{F}$ and $f(e^{i\theta}) \in \partial D$ for almost all $\theta$, then $f \in E\mathscr{F}$ [**12**]. This result is quite different from the earlier facts in that now $E\mathscr{F}$ is a much larger set. If $F$ denotes a conformal mapping of $\Delta$ onto $D$ with $F(0) = w_0$, then $\mathscr{F}$ is described as the set of functions of the form $f = F \circ \phi$, where $\phi \in \mathscr{A}$, $|\phi(z)| < 1$ for $|z| < 1$ and $\phi(0) = 0$. Because of the univalence of $F$, the condition that $f(e^{i\theta}) \in \partial D$ for almost all $\theta$ can be shown to be equivalent to $|\phi(e^{i\theta})| = 1$ for almost all $\theta$. Such functions $\phi$ are called inner functions [**8**, p. 24].

Examples of inner functions are given by finite Blaschke products

$$\phi(z) = x\prod_{k=1}^{n}\frac{z+\alpha_k}{1+\bar{\alpha}_k z},$$

where $|x| = 1$ and $|\alpha_k| < 1$. Since the Möbius transformation $w = (z+\alpha)/(1+\bar{\alpha}z)$ satisfies $|w| = 1$ when $|z| = 1$ (given $|\alpha| < 1$), finite Blaschke products satisfy $|\phi(z)| = 1$ when $|z| = 1$. Convergent infinite Blaschke products also are inner functions.

The converse, that if $f \in E\mathscr{F}$ then $f(e^{i\theta}) \in \partial D$ for almost all $\theta$, is also correct when $D$ is a strip or an angular wedge with opening less than $\pi$. A concise statement of this result is given for a strip. Let $\mathscr{F}$ denote the subset of $\mathscr{A}$ of functions $f$ such that $-1 < \operatorname{Re} f(z) < 1$ for $|z| < 1$ and $f(0) = 0$. A function $f$ in $\mathscr{F}$ is in $E\mathscr{F}$ if and only if $\operatorname{Re} f(e^{i\theta}) = \pm 1$ for almost all $\theta$. This was proved by J. G. Milcetich in [**20**]. The result for wedges was proved by Y. Abu-Muhanna and the author in [**1**]. It is an open problem whether the same result holds when $D$ is any (bounded or unbounded) convex polygonal region.

In the case $D = \Delta$ and $w_0 = 0$, $E\mathscr{F}$ consists of the functions $f$ in $\mathscr{F}$ satisfying

$$\int_0^{2\pi} \log(1 - |f(e^{i\theta})|)\, d\theta = -\infty. \tag{23}$$

This condition is equivalent to the same condition known for the extreme points of the family of functions satisfying $|g(z)| \leqslant 1 (|z| < 1)$ [**8**, p. 125], since the two families are related by the linear homeomorphism defined by $f(z) = zg(z)$ (from $\mathscr{A}$ into $\mathscr{A}_0$). Condition (23) may be regarded as an analytic measure of how close $|f(e^{i\theta})|$ is to 1. It is much less restrictive than $f$ being an inner function. In particular, if $|f(e^{i\theta})| = 1$ on a set of positive measure, then (23) holds.

As an example in the opposite direction, if $f$ is analytic in $\bar{\Delta}$, $|f(z)| \leqslant 1$ for $|z| \leqslant 1$ and $|f(z)| = 1$ only for a finite number of points (on $\partial\Delta$), then $f$ is not an extreme point. (For example, $f(z) = \frac{1}{2}(1+z)$ has these properties.) To verify this claim, it shall be shown that $f$ satisfies $\int_0^{2\pi} \log(1 - |f(e^{i\theta})|)\, d\theta > -\infty$. It suffices to show that if $|f(z_0)| = 1$ and $z_0 = e^{i\alpha_0}$, then there is a closed interval $I$ containing $\alpha_0$ in its interior for which $\int_I \log(1 - |f(e^{i\theta})|)\, d\theta > -\infty$. It may be assumed that $z_0 = 1$. The analyticity of $f$ on $\partial\Delta$ implies that $u(\theta) = \operatorname{Re} f(e^{i\theta})$ and $v(\theta) = \operatorname{Im} f(e^{i\theta})$ are analytic in $\theta$ in a neighborhood of the real axis. The same conclusion holds for the function $|f(e^{i\theta})|^2 = u^2(\theta) + v^2(\theta)$. This function is not identically 1, and thus it has a power series about $\theta = 0$ which begins $1 - c\theta^n + \cdots$, where $c > 0$ and $n$ is an even integer. Thus $|f(e^{i\theta})|$ has a power series beginning $1 - \frac{c}{2}\theta^n + \cdots$, and

$$\log(1 - |f(e^{i\theta})|) \geqslant \log\left(\frac{c}{4}\theta^n\right) = \log\frac{c}{4} + n\log|\theta|$$

for all sufficiently small values of $\theta$. Since $\int_0^1 \log\theta\, d\theta > -\infty$, it follows that $\int_I \log(1 - |f(e^{i\theta})|)\, d\theta > -\infty$ for a sufficiently small closed interval $I$ containing 0 in its interior.

Let $D$ be a bounded, convex region such that $\partial D$ is given by $w = w(t)$, $0 \leqslant t \leqslant 1$, where $w$ has a continuous second derivative. When $f \in \mathscr{F}$, let $\lambda(\theta)$ denote the distance between $f(e^{i\theta})$ and $\partial D$. Assume that the curvature of $\partial D$ is never zero. Then $f \in E\mathscr{F}$ if and only if $f \in \mathscr{F}$ and

$$\int_0^{2\pi} \log\lambda(\theta)\, d\theta = -\infty. \tag{24}$$

This was proved in [**1**]. In the case $D = \Delta$, $\lambda(\theta) = 1 - |f(e^{i\theta})|$. As an example, (24) characterizes the extreme points when the convex region is the interior of an ellipse. This and other results in this section may be roughly summarized by the statement, "the more extreme points $D$ has, the more extreme points $\mathscr{F}$ has."

**9. Meromorphic, univalent functions.** Let $\Sigma$ denote the set of functions that are analytic and univalent in $\{z : 0 < |z| < 1\}$ and have a Laurent series normalized by

$$f(z) = \frac{1}{z} + \sum_{n=0}^{\infty} a_n z^n (0 < |z| < 1). \tag{25}$$

It is preferable to study the subset of $\Sigma$, denoted $\Sigma_0$, for which $a_0 = 0$ in (25), since the functions $f(z) - \frac{1}{z}$, where $f \in \Sigma_0$, can be shown to form a compact subset of $\mathscr{A}$.

The so-called area theorem asserts that if $f \in \Sigma$, then

$$\sum_{n=1}^{\infty} n|a_n|^2 \leqslant 1 \tag{26}$$

[**29**, p. 320]. Inequality (26) is equivalent to the fact that the area (Lebesgue measure) $A$ of the complement of $f[\{z: 0 < |z| < 1\}]$ is nonnegative. For example, the function $f(z) = \frac{1}{z} + z$ belongs to $\Sigma$ and the complement of its range is the interval $\{z: -2 \leqslant x \leqslant 2\}$. For $\Sigma_0$, the equality $|a_1| = 1$ holds only for $\frac{1}{z} + xz$ where $|x| = 1$.

A function $f$ in $\Sigma_0$ belongs to $EH\Sigma_0$ if and only if $A$ is zero. One direction of this result is due to G. Springer [**30**]. The argument depends on showing that the set of functions $g(z) = \sum_{n=1}^{\infty} a_n z^n$ for which (26) holds has as its extreme points those functions for which equality holds in (26). This is analogous to the fact that in $R^n$ the set of extreme points of the closed unit ball consists of the boundary points of that ball. The other direction of the result about $EH\Sigma_0$ was proved by D. H. Hamilton in [**15**]. The argument depends on a result of Nguyen X. Uy [**32**], which gives the construction of a suitable non-constant analytic function defined in the complement of a set of positive measure.

**10. Support points.** Suppose that $\mathscr{F}$ is a compact subset of $\mathscr{A}$ and $f \in \mathscr{F}$. Then $f$ is called a support point of $\mathscr{F}$ if there is a continuous linear functional $L$ such that $\operatorname{Re} L$ is not constant on $\mathscr{F}$ and

$$\operatorname{Re} J(f) = \max\{\operatorname{Re} J(g) : g \in \mathscr{F}\}. \tag{27}$$

In other words, support points are the solutions of nontrivial linear problems over $\mathscr{F}$. The notion of a support point applied to a set in $R^n$ means that certain boundary points of the set are contained in a supporting plane to the set. A general problem, of considerable research interest in the last several years, is that of finding the set of support points of a given subset of $\mathscr{A}$. The set of support points of $\mathscr{F}$ and the set $EH\mathscr{F}$ are related, and a particular connection was indicated in the proof of (16).

Next, some results about the support points of the families discussed earlier are mentioned. A summary of facts about the support points of $S$ is contained in [**9**, pp. 280ff.].

The set of support points of $\mathscr{P}$ consists of all functions of the form

$$p(z) = \sum_{k=1}^{n} t_k \frac{1 + x_k z}{1 - x_k z}, \tag{28}$$

where $|x_k| = 1$, $t_k \geqslant 0$ and $\sum_{k=1}^{n} t_k = 1$ (see [**6**, 13]). This result has the implication that the set of support points of $\mathscr{P}$ is dense in $\mathscr{P}$, a quite different phenomenon than occurs in a finite dimensional setting. A similar result holds for families having ranges in a strip and in a convex angular wedge. A general setting for results of this type concerns families of analytic functions defined by $f(\Delta) \subset D$ and $f(0) = w_0$, where $D$ is a convex domain. If we let $F$ denote a one-to-one analytic mapping of $\Delta$ onto the convex domain $D$ with $F(0) = w_0$, then the support points of $\mathscr{F}$ are characterized by $f = F \circ \phi$, where $\phi$ is a finite Blaschke product and $\phi(0) = 0$ [**13**]. In particular, the set of support points of the family of functions analytic in $\Delta$ and satisfying $|f(z)| \leqslant 1$ are the finite Blaschke products [**6**].

A simple situation occurs for $S^*$, $K$ and the family of functions $f$ in $\mathscr{A}$ satisfying $f(z) \neq 0$,

$|\arg f(z)| < \alpha\frac{\pi}{2}$ and $f(0) = 1$, with $\alpha > 1$. In each the set of support points is the same as $EH\mathscr{F}$ [5], [13].

## References


**1.** Y. Abu-Muhanna and T. H. MacGregor, Extreme points of families of analytic functions subordinate to convex mappings, Math. Z., 176 (1981) 511–519.

**2.** L. V. Ahlfors, Complex Analysis, 3rd edition, McGraw-Hill, New York, 1979.

**3.** R. V. Benson, Euclidean Geometry and Convexity, McGraw-Hill, New York, 1966.

**4.** D. A. Brannan, J. G. Clunie, and W. E. Kirwan, On the coefficient problem for functions of bounded boundary rotation, Ann. Acad. Sci. Fenn. Ser. A I Math., 523 (1973) 18 pp.

**5.** L. Brickman, T. H. MacGregor, and D. R. Wilken, Convex hulls of some classical families of univalent functions, Trans. Amer. Math. Soc., 156 (1971) 91–1

**6.** . C. Cochrane and T. H. MacGregor échet differentiable functionals and support points of families of analytic functions, Trans. Amer. Math. Soc., 236 (1978) 75–92.

**7.** N. Dunford and J. T. Schwartz, Linear Operators. I: General Theory, Pure and Appl. Math., vol. 7, Interscience, New York, 1958.

**8.** P. L. Duren, Theory of $H^p$ Spaces, Academic Press, New York, 1970.

**9.** ______, Univalent Functions, Springer-Verlag, New York, 1983.

**10.** H. G. Eggleston, Convexity, Cambridge Univ. Press, Cambridge, 1963.

**11.** A. W. Goodman, The valence of sums and products, Canad. J. Math., 20 (1968) 1173–1177.

**12.** D. J. Hallenbeck and T. H. MacGregor, Subordination and extreme-point theory, Pacific J. Math., 50 (1974) 455–468.

**13.** ______, Support points of families of analytic functions described by subordination, Trans. Amer. Math. Soc., 278 (1983) 523–546.

**14.** ______, Linear Problems and Convexity Techniques in Geometric Function Theory, Pitman Publishing Co., Boston, 1984.

**15.** D. H. Hamilton, The extreme points of $\Sigma$, Proc. Amer. Math. Soc., 85 (1982) 393–396.

**16.** G. Herglotz, Über Potenzreihen mit positivem, reelem Teil im Einheitskreis, S.-B. Sächs Akad. Wiss. Leipzig Math.-Natur. Kl., 63 (1911) 501–511.

**17.** T. H. MacGregor, Applications of extreme-point theory to univalent functions, Michigan Math. J., 19 (1972) 361–376.

**18.** ______, Hull subordination and extremal problems for starlike and spirallike mappings, Trans. Amer. Math. Soc., 183 (1973) 499–510.

**19.** ______, Variability regions for families of univalent functions, Indiana Univ. Math. J., 31 (1982) 689–699.

**20.** J. G. Milcetich, On the extreme points of some sets of analytic functions, Proc. Amer. Math Soc., 45 (1974) 223–228.

**21.** R. R. Phelps, Lectures on Choquet's Theorem, Van Nostrand, New York, 1966.

**22.** G. Pólya and G. Szegő, Problems and Theorems in Analysis (2 vols.), Springer-Verlag, Berlin, 1972, 1976.

**23.** Chr. Pommerenke, Univalent Functions, Vandenheock and Ruprecht, Göttingen, 1975.

**24.** P. Porcelli, Linear Spaces of Analytic Functions, Rand McNally, Chicago, 1966.

**25.** F. Riesz, Sur certains systèmes singuliers d'équations intégrales, Ann. École Norm. Sup., 28 (1911) 33–62.

**26.** R. M. Robinson, Univalent majorants, Trans. Amer. Math. Soc., 61 (1947) 1–35.

**27.** St. Ruscheweyh, Nichtlineare Extremalprobleme für holomorphe Stieltjesintegrale, Math. Z., 142 (1975) 19–23.

**28.** G. Schober, Univalent Functions-Selected Topics, Lecture Notes in Mathematics no. 478, Springer-Verlag, Berlin, 1975.

**29.** H. Silverman, Complex Variables, Houghton Mifflin, Boston, 1975.

**30.** G. Springer, Extreme Punkte der konvexen Hülle schlichte Funktionen, Math. Ann., 129 (1955) 230–232.

**31.** M. Tsuji, Potential Theory in Modern Function Theory, Maruzen Co., Tokyo, 1975.

**32.** Nguyen Xuan Uy, Removable sets of analytic functions satisfying a Lipschitz condition, Ark. Mat., 17 (1979) 19–27.

**33.** F. A. Valentine, Convex Sets, McGraw-Hill, New York, 1964.


---

### ANSWER TO PHOTO ON PAGE 391

Mark Kac (1914–1984) enjoyed describing himself as only a simple Pole.

# TOWERS OF HANOI AND ANALYSIS OF ALGORITHMS

PAUL CULL AND E. F. ECKLUND, JR.
*Department of Computer Science, Oregon State University, Corvallis, OR 97331*

**1. Introduction.** Mathematics has applications everywhere, but perhaps nowhere more clearly than in computer science, and within computer science perhaps nowhere more clearly than in analysis of algorithms. It would be difficult to draw a line and say that this part of analysis of algorithms is mathematics and that part is computer science.

In this paper we want to give a feeling for analysis of algorithms by investigating, in some detail, algorithms for the Towers of Hanoi problem. But first we should say a few general words about the field of analysis of algorithms. As a start one should make a bow to Donald Knuth. His "Art of Computer Programming" [**5**] is the standard work in the field. In the last few years, analysis of algorithms courses have become a part of many undergraduate programs in computer science. Knuth is still the standard reference, but more widely used textbooks are Aho, Hopcroft and Ullman (AHU) [**1**], Baase [**2**], and Horowitz and Sahni [**4**].

The task of analysis of algorithms is three-fold:

(1) To produce provably correct algorithms, that is, algorithms which not only solve the problem they are designed to solve, but which also can be demonstrated to solve the problem;

(2) To compare algorithms for a problem with respect to various measures of resources (e.g., time and space), so that we can say when one algorithm is better than another;

(3) To find, if possible, the best algorithm for a problem with respect to a particular measure of resource usage. This involves proving "lower bounds", that is, showing that every algorithm which solves the problem must use at least so much of a particular resource. To establish a best algorithm one must have both a proof of a lower bound and an algorithm which uses no more than this lower bound.

Here a distinction should be made between bounds for an algorithm and bounds for a problem. If one establishes an upper bound on a particular resource used by an algorithm for a problem, then one has an upper bound both for the algorithm and for the problem. If one establishes a lower bound for a problem, then one also has a lower bound on all algorithms which solve this problem. But demonstrating a lower bound for one algorithm for a problem does *not* establish a lower bound for the problem.

In this paper we will exemplify the three-fold task of analysis of algorithms using the Towers of Hanoi problem. This problem is often used as an example of a problem which can be neatly solved by a recursive algorithm, as an example of a problem which requires exponential time for its solution [**5**], and as an example of problem solving strategies [**6**]. In the Towers of Hanoi problem one is given three towers, usually called A, B and C, and $n$ disks of different sizes. Initially the disks are stacked on tower A in order of size (disk $n$, the largest, on the bottom; disk 1, the smallest, on the top). The problem is to move the stack of disks to tower C, moving the

*Paul Cull*: I studied at Providence College and the University of Chicago, where I earned my Ph.D. in mathematical biology in 1970. Since then I have been on the faculty at Oregon State University, where I am now professor of computer science. When I am not in Oregon I like to be in Italy, where on various occasions I have been associated with the Laboratory of Cybernetics, near Naples, and the University of Salerno. My main research interests are analysis of algorithms and automata theory. I also continue my interest in mathematical biology, particularly the dynamics of populations and neural nets. I believe that one's teaching and research should interact. The material in this paper grew out of my attempts to devise a good example of an analysis of a problem which could be understood by my classes and did not involve a problem, like sorting, which they had already seen in several other courses.

*Earl F. Ecklund, Jr.*: I received my Ph.D. in mathematics from Washington State University in 1972. Recently I joined the Computer Research Labs of Tektronix, Inc., moving from Oregon State University. My current research interests are database systems and operating systems in a distributed computing system. My avocations include number theory, particularly factoring algorithms, and squash.

disks one at a time in such a way that a disk is never stacked on top of a smaller disk. An extra constraint is that the sequence of moves should be as short as possible. An algorithm solves the Towers of Hanoi problem if, when the algorithm is given as input $n$ the number of disks, and the names of the towers, then the algorithm produces the shortest sequence of moves which conforms to the above rules.

In this paper we will investigate a variety of algorithms which solve the Towers of Hanoi problem. We will prove the correctness of each algorithm, calculate the time and space used by each algorithm to allow a comparison among them, and prove the lower bounds on time and space required by any algorithm which solves the problem. We will show that the final algorithm which attains these bounds is the best possible for these measures.

**2. Comparing algorithms.** For any (solvable) problem there will be an infinity of algorithms which solve the problem. How do we decide which is the "best" algorithm? There are a number of possible ways to compare algorithms. We will concentrate on two measures: time and space. We would like to say that one algorithm is faster, uses less time, than another algorithm if when we run the two algorithms on a computer the faster one will finish first. Unfortunately, to make this a fair test we would have to keep a number of conditions constant. For example, we would have to code the two algorithms in the same programming language, compile the two programs using the same compiler, and run the two programs under the same operating system on the same computer, and have no interference with either program while it is running. Even if we could practically satisfy all these conditions, we might be chagrined to find that algorithm A is faster under conditions C, but that algorithm B is faster under conditions D.

To avoid this unhappy situation we will only calculate time to order. We let $n$ be some measure of the size of the problem, and give the running time as a function of $n$. For example, in the Towers of Hanoi we will use $n$ for the number of disks. We do not distinguish running times of the same order. For our purposes two functions of $n$, $f(n)$ and $g(n)$, have the same order if for some $N$ there are two positive constants $C_1$ and $C_2$ so that

$$C_1|g(n)| \leqslant |f(n)| \leqslant C_2|g(n)| \quad \text{for all } n \geqslant N.$$

We symbolize this relation by $f(n) = \theta(g(n))$, read $f(n)$ is order $g(n)$. Thus we will consider two algorithms to take the same time if their running times have the same order. In particular, we do not distinguish between algorithms whose running times are constant multiples of one another.

If we find that algorithm A has a time order which is strictly less than algorithm B, then we can be confident that for any large enough problem algorithm A will run faster than algorithm B, regardless of the actual conditions. On the other hand if algorithms A and B have the same time order, then we will not predict which one will be faster under a given set of actual conditions.

The space used by an algorithm is the number of bits the algorithm uses to store and manipulate data. We expect the space to be an increasing function of $n$, the size of the problem. This space measurement ignores the number of bits used to specify the algorithm, which has a fixed constant size independent of the size of the problem. Since we have chosen bits as our unit, we can be more exact about space than we can be about time. We can distinguish an algorithm which uses $3n$ bits from an algorithm which uses $2n$ bits. But we will not distinguish an algorithm which uses $3n + 7$ bits from an algorithm which uses $3n + 1$ bits, because we can hide a constant number of bits within the algorithm itself.

So we will say that we have the "best" algorithm for a problem if we can show that the algorithm has minimal time order, and uses minimal space to within an additive constant.

It is not clear that such a best algorithm must exist. In some problems there is a time-space trade-off; a faster algorithm requires more space. We will demonstrate that this sort of trade-off does not exist in the Towers of Hanoi problem by eventually presenting an algorithm which achieves simultaneously minimal time and minimal space.

**3. A recursive algorithm.** The road to a best algorithm starts with some algorithm which one

then attempts to improve. One often uses some sort of strategy to create an algorithm. A very useful strategy is to look at the problem and see if the solution can be expressed in terms of the solutions of several problems of the same kind, but of smaller size. This strategy is usually called divide-and-conquer. If the problem yields to the divide-and-conquer approach, one can construct a recursive algorithm which solves the problem. This construction also gives almost immediately an inductive proof that the algorithm is correct. Time and space analyses of a divide-and-conquer algorithm are often straightforward, since the algorithm directly gives difference equations for time and space usage.

While these divide-and-conquer algorithms have many nice properties, they may not use minimal time and space. They may, however, serve as a starting point for constructing more efficient algorithms.

Consideration of the Towers of Hanoi problem leads to the key observation that moving the largest disk requires that all of the other disks are out of the way. Hence the $n-1$ smaller disks should be moved to tower B, but this is just another Towers of Hanoi problem with fewer disks. After the largest disk has been moved the $n-1$ smaller disks can be moved from B to C; again this is a smaller Towers of Hanoi problem. These observations lead to the following recursive algorithm [**6**], [**7**], [**9**]:

```
PROCEDURE HANOI(A,B,C,n)
  IF n=1 THEN move the top disk from tower A to tower C
         ELSE HANOI(A,C,B,n-1)
              move the top disk from tower A to tower C
              HANOI(B,A,C,n-1).
```

Is this the best algorithm for the problem? We will show that this algorithm has minimum time complexity, but does not have minimum space complexity. First, though we prove that the algorithm correctly solves the problem, the first task of analysis of algorithms as outlined in the Introduction.

PROPOSITION 1. *The recursive algorithm* HANOI *correctly solves the Towers of Hanoi problem.*

*Proof.* Clearly the algorithm gives the correct minimal sequence of moves for 1 disk. If there is more than one disk, the algorithm moves $n-1$ disks to tower B, then moves the largest disk to tower C, and then moves the $n-1$ disks from tower B to tower C. This is precisely what is required in a minimum move algorithm because according to the rules the largest disk can only be moved when all the other $n-1$ disks are on a single tower. So the $n-1$ disks must be moved from tower A to some other tower. Clearly at least one move is required to move the largest disk from tower A to tower C. When the largest disk is moved to tower C, the other $n-1$ disks are on a single tower and still have to be moved to tower C. By inductively assuming $n-1$ disks are moved in the minimum number of moves, we see that the algorithm for $n$ disks makes no more than the minimal number of moves and finishes with all the $n$-disks moved from tower A to tower C. ■

Here we should remark that we have not only produced a provably correct algorithm for the problem; we have also shown that the minimal sequence of moves is unique. This uniqueness makes the proof of correctness easy. The proof would be more complicated if more than one minimum sequence were possible.

We would like to calculate the running time of HANOI, but we don't know how long various operations will take. How long will it take to move a disk? How long will it take to subtract 1 from $n$? How long will it take to test if $n = 1$? How long will it take to issue a procedure call?

Because we only wish to calculate time to order we don't have to answer these questions exactly, but we do have to make a distinction between operations which take a constant amount of time, independent of $n$, and operations whose running time depends on $n$.

One possibility is to assume that each operation takes constant time independent of $n$. AHU [**1**] calls this assumption the uniform cost criterion. With this uniform cost assumption and letting $T(n)$ be the running time for $n$ disks we have the difference equation

$$T(n) = 2T(n-1) + c,$$

because there are 2 calls to the same procedure with $n-1$ disks and $c$ is the sum of the constant running times for the various operations. Letting $T(1)$ be the running time of the algorithm for 1 disk, we find

$$T(n) = (T(1) + c)2^{n-1} - c,$$

which can be verified by direct substitution. This gives

$$T(n) = \theta(2^n)$$

since

$$\frac{T(1)}{2}2^n \leqslant T(n) < \left(\frac{T(1)+c}{2}\right)2^n.$$

Another possibility is to assume that some of the operations have running times which are a function of $n$. But which function of $n$ should we use? Each of the numbers in the algorithm is between 1 and $n$, and the disks can also be represented by numbers between 1 and $n$. Since such numbers can be represented using about $\log n$ bits, it seems reasonable to assume that each operation which manipulates numbers or disks has a running time which is a constant times $\log n$. AHU [**1**] calls this the logarithmic cost criterion and suggests using it when the numbers used by an algorithm do not have fixed bounds. Using the logarithmic cost criterion we have the difference equation

$$T(n) = 2T(n-1) + c\log n$$

for the running time of the algorithm. This difference equation has the solution

$$T(n) = 2^n\left[\frac{T(1)}{2} + c\sum_{i=1}^{n}\frac{\log i}{2^i}\right],$$

which can be verified by substitution. Since the summation in this solution converges, as one can demonstrate by the ratio test, and assuming that the constants are positive, we have

$$T(n) = \theta(2^n).$$

Since both cost criteria give the same running time, we conclude:

PROPOSITION 2. *The algorithm* HANOI *has running time* $\theta(2^n)$.

Although we have established the running time for a particular algorithm which solves the Towers of Hanoi problem, we have not yet established the time complexity of the problem. We need to establish a lower bound so that every algorithm which solves the problem must have running time greater than or equal to the lower bound. We establish $\theta(2^n)$ as the lower bound in the proof of the following proposition.

PROPOSITION 3. *The Towers of Hanoi problem has time complexity* $\theta(2^n)$.

*Proof*. Following the proof of Proposition 1, a straightforward induction shows that the minimal number of moves needed to solve the Towers of Hanoi problem is $2^n - 1$. Since each move requires at least constant time, we have established the lower bound on time complexity.

An upper bound for the time complexity of the problem comes from Proposition 2. Since the

upper bound and lower bound are equal to order, we have established the $\theta(2^n)$ time complexity of the problem. ■

Now that we know HANOI's time complexity we would like to consider its space complexity. First we will establish a lower bound on space which follows from the lower bound on time.

PROPOSITION 4. *Any algorithm which solves the Towers of Hanoi problem must use at least $n$ + constant bits of storage.*

*Proof.* Since the algorithm must produce $2^n - 1$ moves to solve the problem, the algorithm must be able to distinguish $2^n$ different situations. If the algorithm did not distinguish this many situations, then the algorithm would halt in the same number of moves after each of the two nondistinguished situations, which would result in an error in at least one of the cases.

The number of situations distinguished by an algorithm is equal to the number of storage situations times the number of internal situations within the algorithm. Since the algorithm has a fixed finite size, it can have only a constant number of different internal situations. The number of storage situations (states) is 2 to the number of storage bits. Thus $C \cdot 2^{\text{BITS}} \geq 2^n$, and so $\text{BITS} \geq n - \log C = n + \text{constant}$. ■

In order to discuss the space complexity of the recursive algorithm, let us now consider the data structure used. Two possible data structures are the array and the stack. An array is a set of locations indexed by a set of consecutive integers so that the information stored at a location in the array can be referenced by indicating the integer which indexes the location. For example, the information at location I in the array ARRAY would be referenced by ARRAY[I]. A stack is a linearly ordered set of locations in which information can be inserted or deleted only at the beginning of the stack.

The towers could each be represented by an array with $n$ locations, and each location would need at most $\log n$ bits. So an array data structure with $\theta(n \log n)$ bits would suffice. Alternately, each tower could be represented by a stack. Each stack location would need $\log n$ bits, so again this is a $\theta(n \log n)$ bit structure. Actually a savings would be made. Since only $n$ disks have to be represented, the stack structure needs only $n$ locations versus the $3n$ locations used by the array structure. Another possible structure is an array in which the $i$th element holds the name of the tower on which the $i$th disk is located. This structure uses only $\theta(n)$ bits. Yet another possibility is to not represent the towers, but to output the moves in the form FROM __TO__. Thus we could use no storage for the towers.

The recursive algorithm still requires space for its recursive stack. When a recursive algorithm calls itself, the parameters for this new call will take the places of the previous parameters, so these previous parameters are placed on a stack from which they can be recalled when the new call is completed. Also placed on the stack is the return address, the position in the algorithm at which execution of the old call should be resumed. All of this information, the parameters and the return address, for a single call are referred to as a stack frame. At most $n$ stack frames will be active at any time and each frame will use a constant number of bits for the names of the towers and $\log n$ bits for the number of disks. So the recursive algorithm will use $\theta(n \log n)$ bits whether or not the towers are actually represented. We summarize these considerations by the following proposition.

PROPOSITION 5. *The recursive algorithm* HANOI *correctly solves the Towers of Hanoi problem and uses $\theta(2^n)$ time and $\theta(n \log n)$ space.*

The recursive algorithm uses more than minimal space. We are faced with several possibilities:

(1) Minimal space is only a lower bound and is not attainable by any algorithm.

(2) Minimal space can only be achieved by an algorithm which uses more than minimal time.

(3) Some other algorithm attains both minimal time and minimal space. By developing a series of iterative algorithms, we will arrive at an algorithm which uses both minimal time and minimal space.

**4. Some iterative algorithms.** As a first step in obtaining a better algorithm, we will consider an iterative algorithm which simulates the recursive algorithm for $n \geqslant 2$. This algorithm RECURSIVE SIM is similar to an algorithm given by Tenenbaum and Augenstein [7], but we have chosen to explicitly keep track of the stack counter because this will aid us in finding an algorithm using even less space.

```
PROCEDURE RECURSIVE SIM (A,B,C,n)
   I:= 1
   L1[1]:= A; L2[1]:= C; L3[1]:= B
   NUM[1]:= n-1 ; PAR[1]:= 1 ; PAR[0]:= 1
   WHILE I ≥ 1 DO
     IF NUM[I] > 1
        THEN L1[I+1]:= L1[I]
             L2[I+1]:= L3[I]
             L3[I+1]:= L2[I]
             NUM[I+1]:= NUM[I] - 1
             PAR[I+1]:= 1
             I:= I+1
        ELSE MOVE FROM L1[I] TO L3[I]
             WHILE PAR[I] = 2 DO
                   I:= I-1
             IF I ≥ 1 THEN MOVE FROM L1[I] TO L2[I]
                           PAR[I]:= 2
                           TEMP:= L1[I]
                           L1[I]:= L3[I]
                           L3[I]:= L2[I]
                           L2[I]:= TEMP
```

The names of the towers are stored in the three arrays L1, L2, L3; the number of disks in a recursive call is stored in NUM; and the value of PAR indicates whether a call is the first or second of a pair of recursive calls.

RECURSIVE SIM sets up the parameters for the call HANOI (A, C, B, $n-1$). When the last move for this call is made, the arrays will contain the parameters for calls with 1 through $n-2$ disks, where each of these calls will have PAR $= 2$. The arrays will still contain the parameters for the (A, C, B, $n-1$) call with PAR $= 1$. The inner WHILE loop will pop each of the calls with PAR $= 2$, leaving the array counter pointing at the (A, C, B, $n-1$) call. Since I will be 1 at this point the IF condition is satisfied and the MOVE FROM L1[I] TO L2[I] accomplishes the MOVE

FROM A TO C of the recursive algorithm HANOI. The following assignment statements set up the call (B, A, C, $n-1$) with PAR = 2. So when the moves for this call are completed all of the calls in the array will have PAR = 2, and the inner WHILE loop will pop all of these calls setting I to 0. Then the IF condition will be false, so no operations are carried out, and the outer WHILE condition will be false so the algorithm will terminate.

PROPOSITION 6. *The* RECURSIVE SIM *algorithm correctly solves the Towers of Hanoi problem, and uses $\theta(2^n)$ time and $\theta(n \log n)$ space.*

*Proof.* Correctness follows since this algorithm simulates the recursive algorithm which we have proved correct. The major space usage is in the arrays. Since each time I is incremented the corresponding NUM[I] is decremented, and since NUM[I] never falls below 1, there are at most $n-1$ locations ever used in an array. The four arrays L1, L2, L3, and PAR use only a constant amount of space for each element, but NUM must store a number as large as $n-1$ so it uses $\theta(\log n)$ bits for an element. Thus the arrays use $\theta(n \log n)$ bits.

Now we have to argue about time usage. Most of the operations deal with constant-sized operands so these operations will take constant time. The exceptional operations are incrementing, decrementing, assigning, and comparing numbers which may have $\theta(\log n)$ bits. A difference equation for the time is

$$T(n) = 2T(n-1) + C\log n,$$

where $T(n)$ is the time to solve a problem with $n$ disks and $C\log n$ is the time for manipulating the numbers with $\theta(\log n)$ bits. As in the proof of Proposition 1 we have $T(n) = \theta(2^n)$. ∎

Notice that this algorithm does not improve on the recursive algorithm, but study of this form can lead to a saving of space. Storing the array NUM causes the use of $\theta(n \log n)$ space. If we did not have to store NUM, the algorithm would use only $\theta(n)$ space. Do we need to save NUM? NUM is used as a control variable, so it seems necessary. But if we look at NUM[1] + 1, we get $n$. When NUM[I + 1] is set, it is set equal to NUM[I] − 1, but then

$$\begin{aligned} \mathrm{NUM}[\mathrm{I}+1] + \mathrm{I} + 1 &= \mathrm{NUM}[\mathrm{I}] - 1 + \mathrm{I} + 1 \\ &= \mathrm{NUM}[\mathrm{I}] + \mathrm{I} = n. \end{aligned}$$

Thus the information we need about NUM is stored in I and $n$. So if we replace the test on NUM[I] = 1 with a test on I = $n-1$, we can dispense with storing NUM and improve the space complexity from $\theta(n \log n)$ to $\theta(n)$. This replacement does not increase the time complexity of any step in the algorithm, so the time complexity remains $\theta(2^n)$.

Our new procedure is

```
PROCEDURE NEW SIM (A,B,C,n)

  I: = 1

  L1[1]:= A ; L2[1]:= C ; L3[1]:= B

  PAR[1]:= 1  ; PAR[0]:= 1

  WHILE I ≥ 1 DO

    IF I ≠ n-1

      THEN L1[I+1]:= L1[I]

           L2[I+1]:= L3[I]

           L3[I+1]:= L2[I]
```

```
PAR[I]:= 1
I:= I+1
MOVE FROM L1[I] TO L3[I]
WHILE PAR[I] = 2 DO
      I:= I-1
IF I ≥ 1 THEN MOVE FROM L1[I] TO L2[I]
              PAR[I]:= 2
              TEMP:= L1[I]
              L1[I]:= L3[I]
              L3[I]:= L2[I]
              L2[I]:= TEMP
```

From the above observation we have:

PROPOSITION 7. NEW SIM *correctly solves the Towers of Hanoi problem and uses* $\theta(2^n)$ *time and* $\theta(n)$ *space.*

Although we have reached $\theta(n)$ space, we would like to decrease the space even further, hopefully to $n +$ constant bits. If we look at the array PAR, we find that the algorithm scans PAR to find the first element not equal to 2, replaces that element by 2 and then replaces all the previous 2's by 1's. This is analogous to the familiar operation of adding 1 to a binary number, in which we find the first 0, replace it by a 1, and replace all the previous 1's by 0's. So it seems that we can replace the array PAR by a simple counter. The number of bits in the counter will, of course, depend on $n$.

So far this has not resulted in any saving of space. Will there be enough information in the counter to determine from which tower we should move a disk? The affirmative answer will enable us to achieve a minimal space algorithm. To motivate the design of our minimal space algorithm, we will examine the sequence of 31 moves needed to solve the problem with 5 disks. This sequence is shown in Table 1.

Every other move in the solution involves moving disk 1. So if we know which tower contains disk 1, we would know from which tower to move, in alternate moves, but we might not know which tower to move to. When we consider the three towers to be arranged in a circle, we see from Table 1 that disk 1 always moves in a counterclockwise direction when we have an odd number of disks. Similarly disk 1 always moves in a clockwise direction when we have an even number of disks. Thus, by keeping track of the tower which contains disk 1, and whether $n$ is odd or even, we would know how to make every other move.

For the moves which do not involve disk 1, we know that the move involves the two towers which do not contain disk 1. Looking again at Table 1, we see that the odd numbered disks always move in the same direction as disk 1 and the even numbered disks always move in the opposite direction. So knowing the towers involved and whether the disk to be moved is odd or even would allow us to decide which way to move.

Can we determine from a counter whether the disk being moved is odd or even? If we look at the COUNT column of Table 1, we see that the position of the rightmost 0 tells us the number of the disk to be moved. Thus a single counter with $n$ bits is sufficient to solve the Towers of Hanoi problem.

| TOWER 0 | TOWER 1 | TOWER 2 | DECIMAL COUNT | COUNT | DISK | FROM | TO |
|---|---|---|---|---|---|---|---|
| 12345 | - | - | 0 | 00000 | 1 | 0 | 2 |
| 2345 | - | 1 | 1 | 00001 | 2 | 0 | 1 |
| 345 | 2 | 1 | 2 | 00010 | 1 | 2 | 1 |
| 345 | 12 | - | 3 | 00011 | 3 | 0 | 2 |
| 45 | 12 | 3 | 4 | 00100 | 1 | 1 | 0 |
| 145 | 2 | 3 | 5 | 00101 | 2 | 1 | 2 |
| 145 | - | 23 | 6 | 00110 | 1 | 0 | 2 |
| 45 | - | 123 | 7 | 00111 | 4 | 0 | 1 |
| 5 | 4 | 123 | 8 | 01000 | 1 | 2 | 1 |
| 5 | 14 | 23 | 9 | 01001 | 2 | 2 | 0 |
| 25 | 14 | 3 | 10 | 01010 | 1 | 1 | 0 |
| 125 | 4 | 3 | 11 | 01011 | 3 | 2 | 1 |
| 125 | 34 | - | 12 | 01100 | 1 | 0 | 2 |
| 25 | 34 | 1 | 13 | 01101 | 2 | 0 | 1 |
| 5 | 234 | 1 | 14 | 01110 | 1 | 2 | 1 |
| 5 | 1234 | - | 15 | 01111 | 5 | 0 | 2 |
| - | 1234 | 5 | 16 | 10000 | 1 | 1 | 0 |
| 1 | 234 | 5 | 17 | 10001 | 2 | 1 | 2 |
| 1 | 34 | 25 | 18 | 10010 | 1 | 0 | 2 |
| - | 34 | 125 | 19 | 10011 | 3 | 1 | 0 |
| 3 | 4 | 125 | 20 | 10100 | 1 | 2 | 1 |
| 3 | 14 | 25 | 21 | 10101 | 2 | 2 | 0 |
| 23 | 14 | 5 | 22 | 10110 | 1 | 1 | 0 |
| 123 | 4 | 5 | 23 | 10111 | 4 | 1 | 2 |
| 123 | - | 45 | 24 | 11000 | 1 | 0 | 2 |
| 23 | - | 145 | 25 | 11001 | 2 | 0 | 1 |
| 3 | 2 | 145 | 26 | 11010 | 1 | 2 | 1 |
| 3 | 12 | 45 | 27 | 11011 | 3 | 0 | 2 |
| - | 12 | 345 | 28 | 11100 | 1 | 1 | 0 |
| 1 | 2 | 345 | 29 | 11101 | 2 | 1 | 2 |
| 1 | - | 2345 | 30 | 11110 | 1 | 0 | 2 |
| - | - | 12345 | 31 | 11111 | | | |

TABLE 1. Towers of Hanoi Solution for 5 disks.

We use these facts to construct the algorithm which follows.

```
PROCEDURE TOWERS (n)

  T:= 0  (*TOWER NUMBER COMPUTED MODULO 3*)

 COUNT:= 0  (*COUNT HAS n BITS*)

       ⎧ 1 if n is even
  P:= ⎨
       ⎩-1 if n is odd

  WHILE TRUE DO

    MOVE DISK 1 FROM T TO T+P
```

```
T:= T+P
COUNT:= COUNT + 1
IF COUNT = ALL 1's THEN RETURN
IF RIGHTMOST 0 IN COUNT IS IN EVEN POSITION
  THEN MOVE DISK FROM T-P TO T+P
  ELSE MOVE DISK FROM T+P TO T-P
COUNT:= COUNT + 1
ENDWHILE
```

| | n | ... 2 | 1 |
|---|---|---|---|
| COUNT | 0 | ... 0 | 0 |

A picture of the storage used for COUNT.

Notice that it has n bits, and that we have called the rightmost bit position 1. The positions from right to left are then odd, even, odd, even....

REMARKS. We can still improve this algorithm by removing the first COUNT : = COUNT + 1 statement and deleting the rightmost bit of COUNT. This would also require changing the numbering of the bits in COUNT so that the rightmost bit is bit 0. An algorithm similar to our TOWERS has recently been published by T. R. Walsh [**8**].

We have to show that TOWERS correctly solves the Towers of Hanoi problem. We do this by proving that a certain sequence of moves has been accomplished when COUNT contains a number of the form $2^k - 1$, so that when $k = n$, the sequence of moves for HANOI (A, B, C, $n$) has been completed and the procedure will terminate since COUNT contains all 1's.

PROPOSITION 8. *When* COUNT $= 2^k - 1$, *that is* COUNT $= \boxed{00\ldots01\ldots1}$ *with* $k$ 1'*s*, *then*

*if* $k \not\equiv n$ (MOD 2), *the correct moves for* HANOI (A, C, B, $k$) *have been completed and* $T$ *contains* 1 (*which represents* B),

*if* $k \equiv n$ (MOD 2), *the correct moves for* HANOI (A, B, C, $k$) *have been completed and* $T$ *contains* 2 (*which represents* C).

*Proof*. If $k = 1$, the single move $T$ to $T + P$ has been completed, which is A to C if $n$ is odd, and is A to B if $n$ is even, and $T$ contains $T + P$, which is 2 if $n$ is odd, and is 1 if $n$ is even. This agrees with our claim.

Notice that COUNT can only take on the value $2^k - 1$ immediately before the IF...RETURN statement. Assume the moves for either HANOI (A, B, C, $k$), or HANOI (A, C, B, $k$) have been completed. If $k \not\equiv n$ (MOD 2), the next move will be A to C, since by assumption $T$ now

contains 1; if $k$ is odd, the move is $T - P$ to $T + P$, which is $1 - (1)$ to $1 + 1$, which represents A to C, and if $k$ is even, the move is $T + P$ to $T - P$, which is $1 + (-1)$ to $1 - (-1)$, which represents A to C. If $k \equiv n$ (MOD 2), the next move will be A to B, since by assumption $T$ now contains 2; if $k$ is odd, the move is $T - P$ to $T + P$, which is $2 - (-1)$ to $2 + (-1)$, which represents A to B, and if $k$ is even, the move is $T + P$ to $T - P$, which is $2 + 1$ to $2 - 1$, which represents A to B.

Next COUNT will be incremented to $\boxed{0\ldots 010\ldots 0}$, i.e., $k$ trailing 0's. When COUNT $= 2^{k+1} - 1$, the algorithm will have repeated the same sequence of moves as before, since it only "sees" the rightmost information in COUNT, with the difference that $T$ will have started with a different value. The different starting value of $T$ will result in a cyclic permutation of the labels. If $k \not\equiv n$ (MOD 2), then the completed moves will be

$$\begin{gathered}\text{HANOI}\,(\text{A}, \text{C}, \text{B}, k)\\ \text{A to C}\\ \text{HANOI}\,(\text{B}, \text{A}, \text{C}, k)\end{gathered}$$

giving HANOI (A, B, C, $k + 1$) with $k + 1 \equiv n$ (MOD 2), and $T$ will contain 2 (i.e., $1 + 1$). If $k \equiv n$ (MOD 2), then the completed moves will be

$$\begin{gathered}\text{HANOI}\,(\text{A}, \text{B}, \text{C}, k)\\ \text{A to B}\\ \text{HANOI}\,(\text{C}, \text{A}, \text{B}, k)\end{gathered}$$

giving HANOI (A, C, B, $k + 1$) with $k + 1 \not\equiv n$ (MOD 2), and $T$ will contain 1 (i.e., $2 + 2$). ■

PROPOSITION 9. *The algorithm* TOWERS *uses* $\theta(2^n)$ *time and* $n +$ *constant bits of space.*

*Proof.* For space usage, there are $n$ bits in COUNT, and a constant number of bits are used for $T$ and $P$.

For time, the initialization takes $\theta(n)$ and the WHILE loop is iterated $2^n - 1$ times. If each iteration took a constant amount of time, we would have $\theta(2^n)$, but the test and increment instruction on count could take time $\theta(n)$ giving $\theta(n2^n)$. So we have to show that only $\theta(2^n)$ time is used.

If the value in COUNT is even, then incrementing and testing will only require looking at one bit. If the value in COUNT is odd, and (COUNT $- 1)/2$ is even, then the algorithm only looks at 2 bits. In fact, the algorithm will look at $k$ bits in COUNT in $2^{n-k}$ cases. Thus the time used will be $\theta(\sum_{k=1}^{n} k \cdot 2^{n-k}) = \theta(2^n)$, since $\sum_{k=1}^{\infty} k \cdot 2^{-k}$ converges. ■

We summarize these results in the following theorem.

THEOREM. *Any algorithm which solves the Towers of Hanoi problem for n disks must use at least* $\theta(2^n)$ *time and* $n +$ *constant bits of storage. The algorithm* TOWERS *solves the problem and simultaneously uses minimum time and minimum space.*

**5. Summary and conclusion.** The goal of best algorithm has been attained. To attain the goal we started by analyzing the problem in a divide-and-conquer fashion and deriving from this analysis a recursive algorithm which we could prove solved the problem. Next we analyzed the time used by this recursive algorithm and argued that to order this time was best possible since any solution of the problem requires $2^n - 1$ moves.

From the lower bound on time we derived a lower bound of $n$ bits on the space used by any algorithm which solves the problem. A space analysis of the recursive algorithm showed that it used more space than our lower bound, and that the space usage was required for the recursive stack.

To decrease the space usage we built an iterative algorithm which directly simulated the recursive algorithm. Since this was a direct simulation, it used the same amount of space, but we

now could investigate whether all the information being stored was necessary. We found that storing the number of disks on each simulated recursive call was unnecessary. This led to a new iterative algorithm which used only $\theta(n)$ space.

We then noticed that one of the arrays was functioning as a counter, but replacing it by a counter did not decrease the space usage. Next we investigated whether there was sufficient information in the counter to tell us which disk to move and where to move it. We found that we could tell which disk to move, but where to move the disk depended on whether the total number of disks was odd or even.

When we added a variable to keep track of the parity of the number of disks and another variable to keep track of the tower which contained the smallest disk, we found that we had sufficient information to solve the problem, and that we could dispense with the arrays which kept track of the tower names for each simulated recursive call.

At this point we had an algorithm which used $n +$ constant number of bits, which was equal to our lower bound on space. We then showed that the algorithm still used minimal time order, and hence we had obtained a best algorithm.

We may remark that there could be other quite different looking algorithms which solve the problem and use minimal time and space. What we have tried to exemplify is a design methodology which is frequently used in deriving good algorithms. We have chosen the Towers of Hanoi problem because for this problem we could arrive at the goal of best algorithm. For other problems the process may get stuck. We might find a provable algorithm and lower bounds for the problem, but find that there is a gap between the running time of the algorithm and the lower bound, or find a gap between the space usage of the algorithm and the lower bound. Often the next crucial insight, like a disk always moves clockwise or counterclockwise, might not be discovered for many years after an algorithm is created. Alternately a discovered algorithm may be a best algorithm, but an insight is needed to raise the lower bounds for the problem.

In any case, we hope that we have given the reader some feel for analysis of algorithms.

**6. Exercises.** To see that you have understood a technique, it is useful to try to use the technique on similar problems. We give here two more algorithms for the Towers of Hanoi problem. Your task, if you decide to accept it, is to show that these algorithms do in fact solve the problem (i.e., prove that they are correct) and to determine the time and space usage of these algorithms.

```
Exercise 1:

    PROCEDURE HANOI ITERATIVE (A,B,C,n)

      IF n mod 2 = 0 THEN MOVE[1]:= A TO B

                          ELSE MOVE[1]:= A TO C

      K:= 1

      WHILE n > 1 DO

          n:= n-1; K:= 2*K

          IF n mod 2 = 0 THEN MOVE[K]:= A TO B

                                  L1:= C; L2:= A; L3:= B

                              ELSE MOVE[K]:= A TO C

                                  L1:= B; L2:= C; L3:= A

          FOR I:= 1 TO K-1 DO
```

```
CASE MOVE[I] OF

     A TO B : MOVE[K+I]:= L1 TO L2

     A TO C : MOVE[K+I]:= L1 TO L3

     B TO A : MOVE[K+I]:= L2 TO L1

     B TO C : MOVE[K+I]:= L2 TO L3

     C TO A : MOVE[K+I]:= L3 TO L1

     C TO B : MOVE[K+I]:= L3 TO L2
```

*Hints* 1. For correctness you may want to introduce a new variable and prove a statement which says that on each iteration of the WHILE loop the new variable increases (or if you want decreases) and that at the end of each iteration a Hanoi problem whose size depends on the new variable has been solved. You will need to give the tower names for the problem which has been solved. You will also need to specify the value of the new variable.

For space, you should know that the algorithm is storing each move in the array MOVE.

For time, you may want to consider both the uniform and the logarithmic cost measures.

```
Exercise 2:  (Buneman and Levy [3])

      MOVE SMALLEST DISK ONE TOWER CLOCKWISE

      WHILE A DISK (OTHER THAN THE SMALLEST) CAN BE MOVED DO

            MOVE THAT DISK

            MOVE THE SMALLEST DISK ONE TOWER CLOCKWISE

      ENDWHILE
```

*Hints* 2. For correctness, you should be careful since this algorithm only solves the original Towers of Hanoi problem when the number of disks is even. You will probably want to introduce a new variable and prove a statement about the configuration of the disks when the number of moves completed is a specific function of your new variable.

For time and space, the algorithm above is incomplete since it doesn't specify the data structure used to determine if a disk can be moved. You might consider representing each tower by a stack of integers with the integers representing the disks on the tower. Alternately you might consider representing the information by an array DISK, so that DISK[I] contains the name of the tower which contains the Ith largest disk. You may also find it useful to show that the $i$th disk is moved $2^{n-i}$ times.

**References**

1. A. Aho, J. Hopcroft, and J. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1974.

2. S. Baase, Computer Algorithms, Addison-Wesley, Reading, MA, 1978.

3. P. Buneman and L. Levy, The Towers of Hanoi Problem, Information Processing Letters 10, 1980, pp. 243–244.

4. E. Horowitz and S. Sahni, Fundamentals of Computer Algorithms, Computer Science Press, Rockville, MD, 1978.

5. D. Knuth, The Art of Computer Programming, vol. 1, Fundamental Algorithms (2nd ed., 1973), vol. 2, Seminumerical Algorithms (2nd ed., 1981), vol. 3, Search and Sorting (1973), Addison-Wesley, Reading, MA.

6. H. Simon, The Functional Equivalence of Problem Solving Skills, Cognitive Psychology, 1975, pp. 268–288.

7. A. Tenenbaum and M. Augenstein, Data Structures Using PASCAL, Prentice-Hall, Englewood Cliffs, NJ, 1981, pp. 149–154.

8. T. R. Walsh, The Towers of Hanoi Revisited: Moving the Rings by Counting the Moves, Information Processing Letters 15, 1982, pp. 64–67.

9. D. Wood, The Towers of Brahma and Hanoi Revisited, Computer Science Technical Report No. 80-CS-23, McMaster University, 1980.

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## A SHORT PROOF OF STEINHAUS' THEOREM ON SUMMABILITY

JEFF CONNOR

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242*

In this note, we offer a soft (and short!) proof of a result of Steinhaus: a regular matrix summability method cannot sum all sequences of 0's and 1's.

Let $x$ and $y$ be sequences. By a *matrix summability method* we mean a sequence to sequence transformation defined by an infinite matrix $A = (a_{n,k})_{n,k=1}^{\infty}$, an array of complex numbers, and $y = Ax$ when

$$y_n = \sum_{k=1}^{\infty} a_{n,k} x_k \quad \text{for all } n \in N.$$

$A$ is called regular if it maps convergent sequences into convergent sequences and preserves limits, i.e., $\lim_n (Ax)_n = \lim_n x_n$ if $x$ is convergent. Regular matrices are characterized by the following:

THEOREM (Silverman, Toeplitz) [**PS**]. *Let $A = (a_{n,k})_{n,k=1}^{\infty}$ be a matrix summability method. $A$ is regular if and only if*

(1) $\lim\limits_n \sum_{k=1}^{\infty} a_{n,k} = 1,$

(2) $\lim\limits_n a_{n,k} = 0$ *for all $k \in N$, and*

(3) *there is an $M > 0$ such that*
$$\sum_{k=1}^{\infty} |a_{n,k}| < M \quad \textit{for all } n \in N.$$

*$A$ sums $x$* or *$x$ is $A$-summable* if $A$ maps $x$ into a convergent sequence.

We also use the Baire Classification Theorem, a direct ancestor of the Baire Category Theorem:

THEOREM (Baire) [**O**, p. 32]. *Let $X$ be a complete metric space and $(f_n)_{n=1}^{\infty}$ a sequence of continuous complex-valued functions. If $f(x) = \lim_n f_n(x)$ exists for all $x \in X$, then $f$ is continuous except at a set of points of first category.*

Let $\Omega$ denote the sequences of 0's and 1's and define a metric $d$ on $\Omega$ by

$$d(x, y) = \sum_{n=1}^{\infty} |x_n - y_n| \cdot 2^{-n}.$$

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S, L.** The Sacred Beetle and Other Great Essays in Science. Ed: Martin Gardner. Prometheus Books, 1984, xv + 427 pp, $22.95. [ISBN: 0-87975-257-2] A revised and expanded hardcover reprinting of Gardner's 1957 paperback collection Great Essays in Science, including four new essays by Stephen Jay Gould, Lewis Thomas, Carl Sagan, and Isaac Asimov--four distinguished science writers whose writing careers began after the original edition was published. LAS

**General, T(13: 1).** The Ascent of Mathematics. Raymond Coughlin, David E. Zitarelli. McGraw-Hill, 1984, xiii + 609 pp, $27.95. [ISBN: 0-07-013215-1] Based on the premise that "every mathematical technique was created by someone to solve a human problem." Text presents some traditional mathematics (elementary number theory, elementary probability) introduced by discussion of their cultural and/or historical settings. Flexible format. Many exercises. JRG

**Education, S(15-16).** Student Merit Awards. Ed: Leroy Sachs. NCTM, 1984. Middle School, v + 100 pp, $6 (P) [ISBN: 0-87353-215-5]; High School, iv + 137 pp, $9 (P). [ISBN: 0-87353-214-7] Series of self-contained units, each requiring 10 to 30 hours of independent work by students. Each includes explanations, references, requirements and a teachers' guide, and is reproducible. Mathematical level and requirements vary to suit varying student abilities and interests. Topics range from purely practical to traditional enrichment. Quality somewhat uneven but every unit workable and interesting. MW

**Education, S(15-16), P, L.** Computers in Mathematics Education, 1984 Yearbook. Viggo P. Hansen, Marilyn J. Zweng. NCTM, 1984, x + 244 pp, $14.50. [ISBN: 0-87353-210-4] Issues, ideas and sample programs for using computers in teaching mathematics K-12. Twenty-six articles discuss philosophical issues, the computer as a teaching aid, using programming to teach mathematics, and computer diagnosis of student errors. MW

**Education, S(15-16).** Multicultural Mathematics Posters and Activities. NCTM, 1984, $7 (P). Set of 18 posters for junior high classrooms emphasizing interdisciplinary applications of mathematics. Each poster accompanied by teaching aids including reproducible worksheets. Topics include ancient numeration and calculation systems, strategy games, mathematics in art, Pascal's triangle and more. Broad cultural and historical range. MW

**Education, T(16), S, P, L.** Computers in Teaching Mathematics. Peter Kelman, et al. Addison-Wesley, 1983, xi + 308 pp, $14.95 (P). [ISBN: 0-201-10565-9] Good basic book on the subject; should be in every mathematics education library. Emphasizes problem solving approach using programming and computer graphics but neglects simulation in secondary mathematics. Concrete suggestions for incorporating computers into the classroom at all levels. Basic explanation of difference between computer programming and computer science. Resource list of software distributors. MW

**History, L.** Handbook of the Napier Tercentenary Celebration or Modern Instruments and Methods of Calculation. Reprint Ser. for the History of Computing, V. 3. Ed: E.M. Horsburgh. Tomash, 1982, xxxi + 343 pp, $42. [ISBN: 0-938228-10-2] Facsimile of the 1914 Handbook of the Exhibition of Napier Relics and of Books, Instruments, and Devices for Facilitating Calculation. Napier's life, tables, calculating machines, ruled papers, and mathematical models. A guide to the exhibit which celebrated Napier's contributions as well as a survey of calculating instruments invented before WWI. MA

**History, L**.** The Genesis of the Abstract Group Concept. Hans Wussing. Transl: Abe Shenitzer. MIT Pr, 1984, 331 pp, $30. [ISBN: 0-262-23109-3] Long awaited translation of a pioneering work traces the evolution of the abstract group concept from implicit group theoretic modes of thought in geometry and number theory through the concept of a group as a permutation group to the axiomatization of the concept. JRG

**Foundations, T(14-15: 1), S.** Introduction to Elementary Mathematical Logic. Abram Aronovich Stolyar. Transl: Elliot Mendelson. Dover, 1983, vii + 209 pp, $5 (P). [ISBN: 0-486-64561-4] Emphasis on the propositional and predicate calculi; one section treats the application of logic to the analysis and synthesis of digital systems. Exercises throughout (no solutions provided). (1970 MIT Press edition, TR, December 1971.) JRG

**Graph Theory, T*(17), S*, P*, L*.** Graphs and Algorithms. Michel Gondran, Michel Minoux. Transl: Steven Vajda. Ser. in Disc. Math. Wiley, 1984, xix + 650 pp, $64.95. [ISBN: 0-471-10374-8] A translation of the 1979 French text. An excellent overview of graph theory and the combinatorial algorithms which link the abstract results and their practical implications. Has a superb collection of exercises. An outstanding bibliographic resource. CEC

**Discrete Mathematics, T(13-14: 2).** Discrete Mathematics. Kenneth A. Ross, Charles R.B. Wright. Prentice-Hall, 1985, xiv + 610 pp. [ISBN: 0-13-215286-X] A substantial introduction to finite, algebraic, algorithmic, and discrete topics, from recursion and semigroups through graphs and trees to homomorphisms and symmetry groups. Designed for freshmen as a course comparable to calculus, it is gently sloped in sophistication and full of thoughtful aids (chapter highlights, notation summaries, Greek alphabet table, even a dictionary of mathematical jargon like "for every," "suppose," "inspection"). Emphasizes abstract algebra more than most similar books, thereby providing a secure base for developing mathematical maturity. LAS

**Number Theory, T*(17: 1), S, P, L.** Number Theory. W. Narkiewicz. Transl: S. Kanemitsu. World Scientific (US Dist: Heyden & Son), 1983, xii + 371 pp, $21 (P). [ISBN: 9971-950-26-X] A sophisticated introduction to the subject. Methods from algebra, topology, analysis and probability are used liberally. Includes the prime number theorem, Dirichlet's theorem, sieve methods, geometry of numbers, additive number theory, probabilistic number theory and algebraic numbers. Has an ample supply of exercises. CEC

**Number Theory, S(18), P.** Diophantine Equations Over Function Fields. R.C. Mason. London Math. Soc. Lect. Note Ser., No. 96. Cambridge U Pr, 1984, x + 125 pp, $15.95 (P). [ISBN: 0-521-26983-0] This volume gives actual solutions for certain Diophantine equations over function fields. In particular, algorithms are provided for the complete resolution of equations of Thue, hyperelliptic and genus one type. Includes a list of references. CEC

**Number Theory, T(18), S, P.** Lectures on Sieve Methods and Prime Number Theory. Y. Motohashi. Springer-Verlag, 1983, 205 pp, $7.90 (P). [ISBN: 0-387-12281-8] Lectures delivered by the author at the Tata Institute in 1981. Part 1 introduces the Selberg sieve, the combinatorial sieve and the linear sieve. Part 2 applies them to fundamental problems in number theory, namely, zero free regions for $\zeta(s)$ and $L(s,\chi)$, zero-density theorems, and primes in intervals and arithmetic progressions. Includes a list of references. CEC

**Linear Algebra, T**(14), S, L.** Elementary Linear Algebra, Fourth Edition. Howard Anton. Wiley, 1984, xiii + 449 pp, $25.95. [ISBN: 0-471-09890-6] The Fourth Edition of this widely-used text has added a new chapter on complex vector spaces, new material on projections, distances in two- and three-dimensional space and a more conventional notation for transition matrices. (First Edition, TR, March 1973; Extended Review, March 1974; Second Edition, TR, May 1977; Third Edition, TR, October 1981.) CEC

**Functional Analysis, T(18: 1), S, P.** Non-Archimedean Analysis: A Systematic Approach to Rigid Analytic Geometry. S. Bosch, U. Güntzer, R. Remmert. Grund. der math. Wissenschaften, B. 261. Springer-Verlag, 1984, xii + 436 pp, $59. [ISBN: 0-387-12546-9] A systematic development of affinoid and rigid analytic geometry from first principles. Includes an introduction to ultrametric analysis and non-Archimedean valuations. No exercises. Extensive bibliography. CEC

**Optimization, T(16-17: 1, 2), L.** Introductory Optimization Dynamics: Optimal Control with Economics and Management Science Applications. Pierre N.V. Tu. Springer-Verlag, 1984, xiii + 387 pp, $14.50 (P). [ISBN: 0-387-13305-4] Deterministic control theory approached primarily from the point of view of the calculus of variations; assumes linear algebra and calculus. Many examples and applications. Extensive bibliography. No exercises. JRG

**Probability, T(18: 1), P.** Asymptotic Methods in Queuing Theory. A.A. Borovkov. Transl: D. Newton. Wiley, 1984, xi + 292 pp, $47.95. [ISBN: 0-471-90286-1] This book concentrates on the methods of asymptotic analysis of queueing processes: it attempts to make these methods as general as possible and to develop the means to investigate complex systems. The book assumes knowledge of probability theory and the theory of random processes. MT

**Statistics, P.** Frontiers in Statistical Quality Control 2. Ed: H.-J. Lenz, G.B. Wetherill, P.-Th. Wilrich. Physica-Verlag, 1984, 292 pp, $43.60 (P). [ISBN: 3-7908-0306-5] A collection of papers from the Second International Workshop on Quality Control held in Kent in 1983. The work represented in this book is classified under three main headings: sampling inspection for attributes, sampling inspection for variables, and sampling inspection for process control. MT

**Statistics, T(17), P.** Self-Organizing Methods in Modeling: GMDH Type Algorithms. Ed: Stanley J. Farlow. Statistics, V. 54. Dekker, 1984, ix + 350 pp, $55. [ISBN: 0-8247-7161-3] A collection of articles assembled to introduce to English-speaking readers the work of Alexey Ivakhenko on the Group Method of Data Handling (GMDH) algorithm. It is an effort to model by successive improvement of equations using only the given input and output information without reference to the mechanics of

the system being modelled. An evolutionary process in which one retains at each iteration the best set of equations--literally survival of the fittest. AWR

**Statistics, P.** Incomplete Data in Sample Surveys. Ed: William G. Madow, Harold Nisselson, Ingram Olkin. Academic Pr, 1983, $50 each. Volume 1: Report and Case Studies, xxiii + 495 pp [ISBN: 0-12-363901-8]; Volume 2: Theory and Bibliographies, xxv + 579 pp [ISBN: 0-12-363902-6]; Volume 3: Proceedings of the Symposium, xxvi + 413 pp [ISBN: 0-12-363903-4]. The Panel on Incomplete Data was organized in 1977 to make a comprehensive review of the literature on survey incompleteness and to explore ways of improving methods of dealing with it. This is a report of the work done by the Panel. MT

**Computer Programming, S*(13).** The Calculus Companion to Accompany Calculus, Second Edition, by Howard Anton. William H. Barker, James E. Ward. Wiley, 1984. Volume 1, $15.95 (P) [ISBN: 0-471-09230-4]; Volume 2, $12.95 (P) [ISBN: 0-471-88614-9]. A hefty collection of self-contained units which correspond to Anton's chapters. A chatty treatment of algebra and trigonometry review, motivation, word problem strategy and detailed computations which marvelously complements the text. No additional problems or proofs. Focuses on deepening rather than expanding the discussion. MA

**Computer Science, S(18), P.** Lecture Notes in Computer Science-177: Programming Languages and Their Definition. H. Bekič. Springer-Verlag, 1984, xxxii + 254 pp, $13 (P). [ISBN: 0-387-13378-X] Includes unpublished papers and manuscripts of the late Hans Bekič of IBM Vienna. Papers are on denotational semantics and mathematical theory of parallelism. FA

**Computer Science, P.** Lecture Notes in Computer Science-175: P-Functions and Boolean Matrix Factorization. André Thayse. Springer-Verlag, 1984, vii + 248 pp, $13 (P). [ISBN: 0-387-13358-5] Covers research results in algorithm description and implementation. Describes an algebraic representation model that allows for the synthesis of algorithms to be implemented on a control automaton (high-level virtual machine) that is realized as an operational automaton (at a low level). FA

**Applications, P, L*.** Cybernetics: Theory and Applications. Ed: Robert Trappl. Hemisphere, 1983, xii + 455 pp, $80. [ISBN: 0-89116-128-7] An introduction to the field, surveying most of its methods and results together with many chapters describing applications (e.g., to biology, linguistics, artificial intelligence, health care, etc.). A valuable reference. JRG

**Applications (Biology), S(18).** Lecture Notes in Biomathematics-53: Evolutionary Dynamics of Genetic Diversity. G.S. Mani. Springer-Verlag, 1984, vii + 312 pp, $17 (P). [ISBN: 0-387-12903-0] Papers from a symposium on "The Basis of Genetic Diversity." Includes a review of polymorphism, its evolutionary significance and a discussion of variability at the enzyme and DNA level. An ecological-genetic model and a "stress-triggered" model are presented. FA

**Applications (Engineering), T(17), P.** Optimisation Methods in Electronics and Communications. Ed: Kenneth W. Cattermole, John J. O'Reilly. Math. Topics in Telecomm., V. 1. Wiley, 1984, 156 pp, $24.95. [ISBN: 0-471-80765-6] The first volume in an intended series for practicing engineers and graduate students in telecommunications who need more training in mathematics. Focus of Volume One is on calculus of variations. AWR

**Applications (Physics), P.** Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics. David Hestenes, Garret Sobczyk. Fundamental Theories of Physics. D Reidel, 1984, xviii + 314 pp, $58. [ISBN: 90-277-1673-0] Clifford algebra is an infinite-dimensional real algebra designed to express geometric concepts of mathematics and mathematical physics in a unified fashion. The authors refine and develop Clifford algebra through what they call geometric algebra to geometric calculus. The goal is to develop a unified language capable of expressing results of differential and integral calculus on manifolds. Further volumes pursuing the same ideas are planned. PZ

**Applications (Social Science), P.** Coalitions and Collective Action. Ed: Manfred J. Holler. Physica-Verlag, 1984, vi + 350 pp, $47.10 (P). [ISBN: 3-7908-0299-9] A collection of short essays on the theory of how coalitions form and on whether they are stable. Game theory and Arrow's Theorem form basic starting points. A lively introduction notes the tendency of economists to build models assuming perfect information, no collusion, pure competition, actors who do not meet other actors but base decisions on supply and demand curves. "Communication, strategies, perceptions, and feelings do not count since information is perfect and the world is homogeneous." Essays in this book do not take this viewpoint. AWR

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

# 1985 Officers and Committees

Mathematical Association of America
Officers and Committees
April 15, 1985

General Offices: Dolciani Mathematical Center
1529 Eighteenth Street, N.W., Washington, DC 20036
Telephone 202-387-5200

Executive Director: Alfred B. Willcox
Associate Director: Marcia P. Sward
Secretary Emeritus: Henry L. Alder

## Officers

President, Lynn Arthur Steen, St. Olaf College (1985-86)
Past-President, Ivan Niven, University of Oregon (1985)
First Vice-President, Gerald L. Alexanderson, University of Santa Clara (1984-85)
Second Vice-President, Ronald M. Davis, Northern Virginia Community College (1985-86)
Secretary, Kenneth A. Ross, University of Oregon (1984-1989)
Treasurer, Leonard Gillman, University of Texas (1983-87)
Editor, Paul A. Halmos, University of Santa Clara (1981-1986)

## Additional Members of the Board of Governors

Past-Presidents: Richard D. Anderson (1984-88); Dorothy L. Bernstein (1982-86)
Elected Members of the Finance Committee: Lida K. Barrett (1984-87) and Donald L. Kreider (1982-85)
Editor of Mathematics Magazine: Doris Schattschneider, Moravian College (1981-85)
Editor of The College Mathematics Journal: Warren Page (1983-88)

Governors-at-Large:

Jeremy Kilpatrick, University of Georgia (1985-87)
Walter Kaufmann-Buhler, Springer-Verlag, New York (1985-87)
Eleanor Green Jones, Norfolk State University (1983-85)
Andrew C-F Liu, University of Alberta (1983-85)
Mary Harley Jones, Fairfax County Schools (1984-86)
Anthony Ralston, SUNY at Buffalo (1984-86)

Sectional Governors (July 1, 1982-June 30, 1985):

Kansas, Sister Jo Ann Fellin, Benedictine College
Missouri, H. Keith Stumpff, Central Missouri State University
New Jersey, Susan G. Marchand, Kean College of New Jersey
Northeastern, Donald B. Small, Colby College
Ohio, William H. Beyer, University of Akron
Pacific Northwest, Calvin T. Long, Washington State University
Seaway, Violet H. Larney, SUNY at Albany
Southeastern, Ivey C. Gentry, Box 7311, Reynolds Station, Winston-Salem, North Carolina
Southwestern, Allen B. Gray, Jr., Northern Arizona University

Sectional Governors (July 1, 1983-June 30, 1986):

Eastern Pennsylvania-Delaware, Howard Anton, Drexel University
Florida, Donald M. Hill, Florida A & M University
Illinois, John D. Bradburn, Elgin Comm. College
Intermountain, Don H. Tucker, University of Utah
Iowa, Donald V. Meyer, Central College
Louisiana-Mississippi, Carol B. Ottinger, Mississippi University for Women
Maryland-Washington, DC-Virginia, John M. Smith, George Mason University
Michigan, George F. Feeman, Oakland University
North Central, Sylvan Burgstahler, University of Minnesota
Southern California, John W. McGhee, Jr., California State University, Northridge
Texas, James C. Bradford, Abilene Christian University

Sectional Governors (July 1, 1984-June 30, 1987):

Allegheny Mountain, Franz X. Hiergeist, West Virginia University
Indiana, Rodney T. Hood, Franklin College
Kentucky, Joe K. Smith, Northern Kentucky University
Metropolitan New York, John A. Thorpe, 950 25th St. NW, Apt. 112N, Washington, DC
Nebraska, David L. Skoug, University of Nebraska, Lincoln
Northern California, Roy W. Ryden, Humboldt State University
Oklahoma-Arkansas, Robert C. Eslinger, Hendrix College
Rocky Mountain, Gary W. Grefsrud, Fort Lewis College
Wisconsin, Paul J. Campbell, Beloit College

Sectional Governors (July 1, 1985-June 30, 1988):

Kansas, Frank S. Brenneman, Tabor College
Missouri, Harold W. Hager, Southeast Missouri State University
New Jersey, Susan G. Marchand, Kean College of New Jersey
Northeastern, James E. Ward III, Bowdoin College
Ohio, Andrew Sterrett, Jr., Denison University
Pacific Northwest, D. George McRae, University of Montana, Missoula
Seaway, Jack E. Graver, Syracuse University
Southeastern, John W. Kenelly, Clemson University
Southwestern, David R. Arterburn, New Mexico Institute of Mining & Tech

## MAA Section Officers

Allegheny Mountain Section:

Chairman: Barbara T. Faires, Westminster College

First Vice-Chairman: W. E. Deskins, University of Pittsburgh
Second Vice-Chairman: James J. Reynolds, Penn St U-Beaver Campus
Coord. of Student Programs: Ronald E. Harrell, Allegheny College
AHSME Reg. Coord. W PA: Nicholas Ford, Penn St U-Fayette Campus
AHSME Reg. Coord. WV: I. D. Peters, Univ. of West Virginia
Newsletter Editor: Richard Reese, Edinboro Univ. of PA
Public Info. Officer: Kathleen Taylor, Duquesne University
Secretary/Treasurer: David T. Brown, Bethany College

Eastern Pennsylvania & Delaware Section:

Chairman: David E. Flesner, Gettysburg College
Vice-Chairman: Bernard Kolman, Drexel University
AHSME Reg. Coord. E PA: Albert E. Filano, West Chester U. of PA
AHSME Reg. Coord. DE: Willard E. Baxter, University of Delaware
Newsletter Editor: Paul Cochrane, Bloomsburg University
Public Info. Officer: Paul Cochrane, Bloomsburg University
Secretary/Treasurer: James O. Brooks, Villanova University

Florida Section:

President: David L. Sherry, University of West Florida
President-Elect: Elizabeth A. Magarian, Stetson University
Past-President: Don R. Lichtenberg, University of South Florida
Vice-President for Programs: Fredric J. Zerla, University of South Florida
Vice-Pres.-Elect. Programs: Kathleen M. Timmer, Jacksonville University
Vice-Pres. for Site Select.: Bill E. Jordan, Seminole Community College
Vice-Pres.-Elect Site Sel.: Michael J. Mears, Manatee Junior College
AHSME Reg. Coord. FL: Roy Bolduc, University of Florida
AHSME Reg. Coord. CZ: Walter E. Mientka, University of Nebraska
AHSME Reg. Coord. PR: Eugene A. Francis, University of Puerto Rico
Newsletter Editor: Robert Kalin, Florida State University
Public Info. Officer: Robert Kalin, Florida State University
Secretary/Treasurer: Ralph D. McWilliams, Florida State University

Illinois Section:

Chairman: John W. Hooker, Southern Illinois University
Chairman-Elect: Linda R. Sons, Northern Illinois University
Past-Chairman: Ronald M. Shelton, Milliken University
First Vice-Chairman: Richard J. Wilders, North Central College
Second Vice-Chairman: Neale R. Fadden, Belleville Area College
AHSME Reg. Coord. IL: Gary Tippett, Bradley University
Secretary/Treasurer: Howard C. Saar, Plainfield Public Schools

Indiana Section:

Chairman: Daniel P. Maki, Indiana University
Vice-Chairman: Catherine M. Murphy, Purdue Univ.-Calumet
AHSME Reg. Coord. IN: Louis J. Cote, Purdue University
Newsletter Editor: G. E. Graves, Rose-Hulman Inst. of Tech
Public Info. Officer: Roger Lautzenheiser, Rose-Hulman Inst. of Tech
Secretary/Treasurer: Duane E. Deal, Ball State University

Intermountain Section:

Chairman: Linda C. Hill, Idaho State University
Past Chairman: Donald W. Robinson, Brigham Young University
First Vice-Chairman: James W. Cotts, So. Utah State College
Second Vice-Chairman: Patricia P. Henry, Weber State College
AHSME Reg. Coord. UT: Ralph V. Skarda, Brigham Young University
AHSME Reg. Coord. ID: Tom Williams, Idaho State University
Public Info. Officer: Linda C. Hill, Idaho State University
Secretary/Treasurer: Lawrence O. Cannon, Utah State University

Iowa Section:

Chairman: Wayne L. Woodworth, Drake University
Chairman-Elect: George N. Trytten, Luther College
AHSME Reg. Coord. IA: Robert M. Buckingham, Lutheran Mutual Life Ins. Co.
Public Info. Officer: George N. Trytten, Luther College
Secretary/Treasurer: Alan J. Heckenbach, Iowa State University

Kansas Section:

Chairperson: Richard S. Rempel, Bethel College
Chairperson-Elect: Kenneth S. Carman, Kansas Wesleyan University
Assoc. Chpsn. Comm. Coll.: Jean Moran, Donnelly College
AHSME Reg. Coord. KS: Robert H. Thompson, Washburn Univ. of Topeka
Public Info. Officer: Elaine Tatham, ETC Institute
Secretary/Treasurer: Robert H. Thompson, Washburn Univ. of Topeka

Kentucky Section:

Chairman: Peter E. Moore, Northern Kentucky Univ.
Chairman-Elect: Rodger Hammons, Morehead State University
Vice-Chairman: Paul E. Bland, Eastern Kentucky Univ.
AHSME Reg. Coord. KY: Carrie C. Parker, Jefferson Comm. Coll. S.W.
Newsletter Editor: Glenn E. Johnston, Morehead State University
Public Info. Officer: Glenn E. Johnston, Morehead State University
Secretary/Treasurer: Sandra A. Spears, Jefferson Community College

Louisiana & Mississippi Section:

Chairman: Stephen A. Doblin, Univ. of Southern Mississippi
Past Chairman: Duane D. Blumberg, U. of Southwestern Louisiana
LA Vice-Chairman: W. T. Thompson, Northeast Louisiana Univ.
MS Vice-Chairman: Rose E. Strahan, Delta State University
AHSME Reg. Coord. LA: James E. Keisler, Louisiana State University
AHSME Reg. Coord. MS: David E. Cook, University of Mississippi
Newsletter Editor: John L. Tilley, Mississippi State University
Public Info. Officer LA: W. T. Thompson, Northeast Louisiana Univ.
Public Info. Officer MS: Rose E. Strahan, Delta State University
Secretary/Treasurer: John L. Tilley, Mississippi State University

Maryland - D.C. - Virginia Section:

Chairman: Howard L. Penn, U.S. Naval Academy
Chairman-Elect: Robert E. Lewand, Goucher College
Vice-Chairman for Programs: Elizabeth J. Teles, Montgomery College
Vice-Chairman for Membership: William M. Sanders, James Madison University
AHSME Reg. Coord. MD-DC-VA: Edward D. Bender, J. Sargeant Reynolds Com Col
Newsletter Editor: William M. Sanders, James Madison University
Public Info. Officer: Caren L. Diefenderfer, Hollins College
Secretary: Caren L. Diefenderfer, Hollins College
Treasurer: V. R. Hancock, Emory and Henry College

Metro. New York Section:

Chairman: Warren Page, N Y City Technical College
Chairman-Elect: Alan Chutsky, Queensborough Comm Coll
Vice-Chairman 4-Year College: F. V. Pohle, Adelphi University
Vice-Chairman 2-Year College: Geoffrey R. Akst, Bor. of Manhattan Comm Coll
Vice-Chairman High School: Sheila Krilov, Hunter College High School
AHSME Reg. Coord. Met NY: Sandra M. Pulver, Pace University
Newsletter Editor: Alan Hoenig, John Jay Col of Crim. Justice
Public Info. Officer: Bernard Sohmer, The City College
Secretary: Lily E. Christ, John Jay Col of Crim. Justice
Treasurer: Robert G. Payton, Adelphi University

Michigan Section:

Chairman: Donald R. Lick, Western Michigan University
Vice-Chairman: Michael J. Gilpin, Michigan Tech University
Vice-Chairman 2-Year College: Barbara K. Near, Henry Ford Comm. College
AHSME Reg. Coord. MI: James Bidwell, Central Michigan University
Newsletter Editor: John O. Kiltinen, Northern Michigan University
Public Info. Officer: Thomas Elsner, General Motors Institute
Secretary/Treasurer: Douglas W. Nance, Central Michigan University

Missouri Section:

President: Larry A. Cammack, Central Missouri University
Past President: Victor Gummersheimer, Southeast Missouri St Univ.
Vice-President: Shirley Huffman, Southwest Missouri St Univ.
Chairman of H.S. Lecturers: Robert E. Kennedy, Central Missouri St Univ.
AHSME Reg. Coord. MO: Alvin R. Tinsley, Central Missouri St Univ.
Newsletter Editor: William A. Ettling, Southeast Missouri St Univ.
Secretary/Treasurer: David W. Bahnemann, Northwest Missouri St Univ.

Nebraska Section:

Chairman: Dale W. Behrens, Hastings College
Chairman-Elect: Alexander Mehaffey Jr., University of South Dakota
AHSME Reg. Coord. NE: Richard L. Barlow, Kearney State College
AHSME Reg. Coord. SD: Jan Vandever, South Dakota St. University
Newsletter Editor: Henry M. Cox, University of Nebraska
Newsletter Editor: Thomas S. Shores, University of Nebraska
Secretary/Treasurer: Henry M. Cox, University of Nebraska

New Jersey Section:

Chairman: S. A. Foote, Rutgers University
Chairman-Elect: Stephanie Sloyan, Georgian Court College
Vice-Chairman Innovation: Thomas J. Marlowe, Seton Hall University
Vice-Chairman Speakers: Patricia K. Kenschaft, Montclair State College
Vice-Chairman 2-Year College: Ruth D. O'Dell, County College of Morris
Vice-Chairman High School: John Sooy, Glassboro State College
AHSME Reg. Coord. NJ: Robert Mitchell, Glassboro State College
Newsletter Editor: F. E. Clark, Rutgers University
Public Info. Officer: F. E. Clark, Rutgers University
Secretary: F. E. Clark, Rutgers University
Treasurer: Catherine Folio, Brookdale Comm. College

North Central Section:

President: Loren C. Larson, St. Olaf College
President-Elect: Lisl N. Gaal, University of Minnesota
Past President: Murray Braden, Macalester College
Member-At-Large: Mary B. Johnson, Inver Hills Comm. College
Member-At-Large: Ronald M. Mathsen, North Dakota St. Univ.
AHSME Reg. Coord. MN: Wayne Roberts, Macalester College
AHSME Reg. Coord. ND: John Whitcomb, University of North Dakota
AHSME Reg. Coord. MB: John Barsby, St. John's Ravenscourt

AHSME Reg. Coord. SK: Don Kapoor, University of Regina
AHSME Reg. Coord. SD: Jan Vandever, South Dakota St. University
Newsletter Editor: Gerald E. Lenz, St. John's University
Public Info. Officer: Joe Gallian, University of Minnesota
Secretary/Treasurer: Stephen R. Hilding, Gustavus Adolphus College

Northeastern Section:

Chairman: Thurmon Whitley, University of New Haven
Vice-Chairman: Steven K. Ingram, Norwich University
2-Year Coll. Representative: Jean B. Smith, Middlesex Community College
AHSME Reg. Coord. CT: James E. Trimble, Hartford Insurance Group
AHSME Reg. Coord. ME: Larry Miller, Union Mutual Life Ins. Co.
AHSME Reg. Coord. MA: Murray Abramson, Bridgewater State College
AHSME Reg. Coord. NH: Owen R. Koppang, Concord High School
AHSME Reg. Coord. RI: Thomasine Manickas, South Kingston H.S.
AHSME Reg. Coord. VT: J.E.H. Johansson, University of Vermont
AHSME Reg. Coord. NB NF NS PE: William Crawford, Mt. Allison University
Newsletter Editor: Philip H. Mahler, Middlesex Community College
Public Info. Officer: Kenneth D. Lane, Colby College
Secretary/Treasurer: Gordon D. Prichett, Babson College

Northern California Section:

Chairman: Howard J. Weiner, University of California
Vice-Chairman: Roseanna F. Torretto, State of California
Program Chairman: Michael Thibodeaux, Menlo College
AHSME Reg. Coord. No CA & NV: Robert McFarland, San Pablo, CA
AHSME Reg. Coord. HI PI: Jack Johnson, Brigham Young U.-Hawaii
AHSME Reg. Coord. GU: Gail S. Mullen, University of Guam
Newsletter Editor: William G. Chinn, City Coll of San Francisco
Public Info. Officer: William G. Chinn, City Coll of San Francisco
Secretary/Treasurer: Leonard Klosinski, University of Santa Clara

Ohio Section:

Chairman: James R. C. Leitzel, Ohio State University
Chairman-Elect: Alan G. Poorman, Ashland College
Past Chairman: Richard A. Little, Baldwin-Wallace
Program Chairman: Robert M. Dieffenbach, Miami University
2-Year College Representative: Cynthia P. Yang, Miami University
AHSME Reg. Coord. OH: Leo Schneider, John Carroll University
AHSME Reg. Coord. WV: I. D. Peters, University of West Virginia
Newsletter Editor: David E. Kullman, Miami University
Public Info. Officer: Susan G. Schneider, John Carroll University
Secretary/Treasurer: Andrew Sterrett Jr., Denison University

Oklahoma-Arkansas Section:

Chairman: John W. Watson, Arkansas Tech University
Past Chairman: Morris L. Marx, University of Oklahoma
First Vice-Chairman: William A. Coberly, University of Tulsa
Second Vice-Chairman: Edward Masley, Arkansas College
AHSME Reg. Coord. OK: Thomas Cairns, University of Tulsa
AHSME Reg. Coord. AR: John Hodges, University of Arkansas
Newsletter Editor: Allan C. Cochran, University of Arkansas
Public Info. Officer: James E. Scroggs, University of Arkansas
Secretary/Treasurer: John M. Jobe, Oklahoma State University

Pacific Northwest Section:

Chairman: Marjorie M. Enneking, Portland State University
Vice-Chairman 2-Year College: Vernon R. Hood, Portland Community College
Vice-Chairman 4-Year College: Harvey J. Schmidt Jr., Lewis & Clark College
Program Chairman: Alan Hoffer, University of Oregon
AHSME Reg. Coord. AK: Donval R. Simpson, Fairbanks, AK
AHSME Reg. Coord. ID: Tom Williams, Idaho State University
AHSME Reg. Coord. OR: Marion Walter, University of Oregon
AHSME Reg. Coord. WA: Calvin Long, Washington State University
AHSME Reg. Coord. AB: G. W. Sands, University of Calgary
AHSME Reg. Coord. BC YK NW: Leon Bowden, University of Victoria
AHSME Reg. Coord. MT: William Myers, University of Montana
Newsletter Editor: Sue Kaplan, Western Washington University
Public Info. Officer: Sue Kaplan, Western Washington University
Secretary/Treasurer: Sam L. Dunn, Seattle Pacific University

Rocky Mountain Section:

Chairman: Ruth R. Struik, University of Colorado
Chairman-Elect: Stephen D. Bronn, Univ. of Southern Colorado
Vice-Chairman: Marie A. Ritten, National College
Program Chairman: Allan G. Skillman, Casper College
AHSME Reg. Coord. CO WY: Bob Vunovich, Univ. of Southern Colorado
AHSME Reg. Coord. MT: William Myers, University of Montana
Newsletter Editor: David W. Ballew, S.D. School of Mines & Tech
Public Info. Officer: David W. Ballew, S.D. School of Mines & Tech
Secretary/Treasurer: David W. Ballew, S.D. School of Mines & Tech

Seaway Section:

Chairperson: Charles L. Smith, SUNY College at Potsdam
Chairperson-Elect: Kenneth Wooster, SUNY College at Cortland
Past Chairperson: Kenneth D. Magill, CUNY Center at Buffalo
First Vice-Chairperson: Rebecca E. Hill, Rochester Institute of Tech
Second Vice-Chairperson: Morton Goldberg, Broome Community College
AHSME Reg. Coord. No NY: Stephen West, SUNY College at Geneseo
AHSME Reg. Coord. W NY: Louis F. Scholl, Trocaire College
AHSME Reg. Coord. ON: F. A. Sherk, University of Toronto
AHSME Reg. Coord. PQ English: Saverio Mirarchi, Laval Catholic High School
AHSME Reg. Coord. PQ French: Paul Charlebois, CEGEP
Newsletter Editor: Donald Trasher, SUNY College at Geneseo
Public Info. Officer: Donald Trasher, SUNY College at Geneseo
Secretary/Treasurer: Donald Trasher, SUNY College at Geneseo

Southeastern Section:

Chairman: F. L. Cook, University of Alabama
Chairman-Elect: Edmond D. Dixon, Tennessee Tech University
Vice-Chairman: Sharon C. Ross, DeKalb Comm. College
Lecturer: John W. Kennelly, Clemson University
AHSME Reg. Coord. AL: James O'Neal, Troy State University
AHSME Reg. Coord. GA: Jan L. Boal, Georgia State University
AHSME Reg. Coord. NC: William Herbst, Jefferson Std. Life Insurance
AHSME Reg. Coord. SC: Stanley Lukawecki, Clemson University
AHSME Reg. Coord. TN: David Collings, Nat'l Life & Acci't Ins. Co.
Newsletter Editor: David R. Stone, Georgia Southern College
Public Info. Officer: David R. Stone, Georgia Southern College
Secretary/Treasurer: James G. Ware, U of Tennessee at Chattanooga

Southern California Section:

Chairman: Donald Y. Goldberg, Occidental College
Past Chairman: Robert F. Brown, UCLA
First Vice-Chairman: David D. Shochat, Logicon
Second Vice-Chairman: Adil M. Yaqub, University of California
Program Chairman: Melvin Henrikson, Harvey Mudd College
Program Vice-Chairman: Herbert A. Gindler, San Diego State University
AHSME Reg. Coord. So CA: Alice M. King, Cal State Polytechnic Univ.
Newsletter Editor: William Margulies, Cal State Long Beach
Public Info. Officer: William Margulies, Cal State Long Beach
Secretary/Treasurer: William Margulies, Cal State Long Beach

Southwestern Section:

Chairman: William Y. Velez, University of Arizona
Vice-Chairman: Richard M. Grassl, University of New Mexico
AHSME Reg. Coord. AZ-NM: Elias Toubassi, University of Arizona
AHSME Reg. Coord. TX: James R. Boone, Texas A&M University
Public Info. Officer: Alvin Swimmer, Arizona State University
Secretary/Treasurer: Edward D. Gaughan, New Mexico State University

Texas Section:

Chair: Margaret R. Hutchinson, University of St. Thomas
Chair-Elect: R. V. Morgan, Sul Ross State University
Immediate Past Chair: Robert H. Cranford, University of Texas
Level I Director: George Brewer, Odessa College
Level II Director: Darwin Peake, Trinity University
Level III Director: James N. Younglove, U. of Houston-Central Campus
Director-At-Large: Rose M. Smith, Texas Woman's University
Arrangements Chair: Joseph Szucs, Texas A&M University-Box 1576
Arrangements Chair: Anthony Blozinski, Texas A&M University-Box 1576
Arrangements Chair-Elect: Vivian A. Dennis, Eastfield College
AHSME Reg. Coord. TX: James R. Boone, Texas A&M University
Newsletter Editor: Glen E. Mattingly, Sam Houston State University
Public Info. Officer: James C. Bradford, Abilene Christian University
Secretary/Treasurer: Glen E. Mattingly, Sam Houston State University

Wisconsin Section:

Chairman: Gary L. Britton, U. of Wisconsin Center
Chairman-Elect: Richard J. Mihalek, University of Wisconsin
Past Chairman: Lois B. Brualdi, Edgewood College
AHSME Reg. Coord. WI: James Muhich, University of Wisconsin
Newsletter Editor: Nicholas C. Bystrom, Northland College
Public Info. Officer: Karl E. Byleen, Marquette University
Secretary/Treasurer: James M. Sobota, University of Wisconsin

## Committees of the Association

Terms of members expire, except where otherwise noted, at the end of the Annual Meeting in January following the last year of service listed below. For temporary committees, no terms are listed since they are automatically discharged at the expiration of the President's term of office, which is the Annual Meeting in January, 1986.

Executive Committee:

Lynn A. Steen, Chairman (1985-87); Gerald L. Alexanderson (1984-85); Ronald M. Davis (1985-86); Leonard Gillman (1983-87); Paul R. Halmos (1982-86); Ivan Niven (1985); Kenneth A. Ross (1984-1989).

Finance Committee:

Lynn A. Steen, Chairman (1985-87), ex-officio; Lida K. Barrett (1984-87); Leonard Gillman (1983-87), ex-officio; Donald L. Kreider (1978-85); Ivan Niven (1983-86), ex-officio; Kenneth A. Ross (1984-1989) ex-officio.

Audit and Budget Committee: Donald L. Kreider, Chairman (1978-85); Lida K. Barrett (1984-87).

Investment Committee: Leonard Gillman, Chairman (1983-87), ex-officio; Henry L. Alder (1984-86); Harley Flanders (1984-86); Gerald J. Porter (1983-85); Kenneth A. Ross (1984-1989) ex-officio; James M. Vaughn, Jr. (1984-86).

Staff and Services Committee: Leonard Gillman, Chairman (1983-87); Lynn A. Steen (1986), both ex-officio.

Advisory Committee on the Archives of American Mathematics:

Judith V. Grabiner, Chairman (1979-85); Churchill Eisenhart (1984-86); Leonard Gillman (1979-85).

Committee on Archival Holdings in Mathematics:

Sanford L. Segal, Chairman (1984-85); Jeanne LaDuke (1984-85); Albert C. Lewis (1984-85); Uta C. Merzbach (1984-85).

Nominating Committee:

William G. Chinn, Chairman (1985); Richard D. Anderson (1985); Jerome A. Goldstein (1985-86); Warren Page (1985-86); Alan C. Tucker (1985-86).

Nominating Committee for the Editor of the Monthly:

Alan C. Tucker, Chairman; Donald J. Albers; Stephen B. Maurer; Barbara L. Osofsky.

Committee on Committees:

Lynn A. Steen, Chairman (1985-87), ex-officio; Ivan Niven (1982-85), ex-officio; Kenneth A. Ross (1985-89), ex-officio; Alfred B. Willcox, ex-officio.

Committee on Corporate Members:

Richard A. Moynihan, Chairman (1981-85); Jeremiah J. Lyons (1981-86); Henry O. Pollak (1983-85); Peter L. Renz (1984-86); Alfred B. Willcox (1980-85), ex-officio.

Committee on Earle Raymond Hedrick Lectures:

James W. Cannon, Chairman (1985-86); Neil J.A. Sloane (1985-88); Elias M. Stein (1984-87).

Committee on American Mathematics Competition (CAMC):

Stephen B. Maurer, Chairman (1981-86); Richard D. Anderson (1985-86); Joyce Becker (1984-85); George Berzsenyi (1983-85); James D. Bristol (1983-85); Bruce Brombacher (1987); Thomas R. Butts (1982-86); Michael Ecker (1985-87); Elgin Johnston (1985-87); Irwin W. Kaufman (1982-85); Thomas R. Knapp (1987); Joseph Konhauser (1982-85); Gerald E. Martau (1984-86); Steven Meiring (1984-86); Walter E. Mientka (1979-85); Robert Musen (1987); Harold Reiter (1984-86); Ian Richards (1987); Stephen Rodi (1987); Leo J. Schneider (1980-85); Margaret W. Tiller (1982-85); Dorothy Wendt (1984-86); Melvin R. Woodard (1981-86).

Subcommittee on the American High School Mathematics Examination: Stephen B. Maurer, Chairman (1981-86); George Berzsenyi (1982-85), ex-officio; Thomas R. Butts (1980-86), ex-officio; Thomas R. Knapp (1987); Gerald E. Martau (1984-86); Walter E. Mientka (1980-85), ex-officio; Harold Reiter (1984-86); Ian Richards (1987); Leo J. Schneider (1980-85); Dorothy Wendt (1984-86).

Subcommittee on the American Junior High School Mathematics Examination: Thomas R. Butts, Chairman (1982-86); Richard D. Anderson (1985-86); Michael Ecker (1985-87); Elgin Johnston (1985-87); Irwin Kaufman (1984-85); Stephen B. Maurer (1981-86), ex-officio; Steven Meiring (1984-86); Walter E. Mientka (1981-85), ex-officio; Stephen Rodi (1987); Melvin R. Woodard (1986).

Subcommittee on the American Invitational Mathematics Examination: George Berzsenyi, Chairman (1982-85); James D. Bristol (1982-85); Michael W. Ecker (1987); Elgin Johnston (1987); Stephen B. Maurer (1981-86), ex-officio; Walter E. Mientka (1980-85), ex-officio; Ian Richards (1987).

Advisory Panel of the Committee on High School Contests: David Bernstein (1985); W.J. Blundon (1985); Kenneth A. Brown, Jr. (1985); Pamela Drummond (1985); Richard A. Gibbs (1985); Samuel L. Greitzer (1985); Douglas A. Hensley (1985); Sheila Krilov (1985); Eric Lander (1985); Eugene McGovern (1985); Richard Parris (1985); Stanley Rabinowitz (1985); Richard Rhoad (1985); Mark Saul (1985); Akehiko Takahashi (1985); Claire Tuckman (1985).

Subcommittee on the USA Mathematical Olympiad: J. Ian Richards, Chairman (1985-87); George Berzsenyi (1980-85), ex-officio; Irwin W. Kaufman (1982-85); Murray S. Klamkin, Problems Chairman (1982-85); Joseph Konhauser (1982-85); Andrew Liu (1981-86); Stephen B. Maurer (1986), ex-officio; Walter E. Mientka (1985), ex-officio; Cecil Rousseau (1985-87).

IMO Training Session: Cecil Rousseau, Director (1985); Gregg Patruno, Assistant Director (1985).

Committee on Placement Examinations:

John W. Kenelly, Chairman (1985); Linda H. Boyd (1984-86); Judith N. Cederberg (1985-87); John G. Harvey (1984-86); Mary McCammon (1985-87); Billy E. Rhoades (1982-85).

Committee on Publications:

Alan C. Tucker, Chairman (1982-85); Donald J. Albers (1985); Duane W. Bailey (1984-86); Charles A. Cable (1982-85); James P. Crawford (1984-86); Ross L. Finney (1980-85); Leonard Gillman (1983-87), ex-officio; Paul R. Halmos (1982-86), ex-officio; Pete Hayslett (1985-87); John Karlof (1982-85); William F. Lucas (1982-87); Donald J. McCarthy (1984-86); Warren Page

(1985-88), ex-officio; Doris Schattschneider (1981-85), ex-officio.

Subcommittee on Carl B. Allendoerfer Awards: JoAnne Growney, Chairman (1985); Underwood Dudley (1985-87); Doris Schattschneider (1982-85), ex-officio; Marjorie Senechal (1986).

Subcommittee on Basic Library Lists: Deborah T. Haimo (1983-86); William F. Lucas (1982-85).

Subcommittee on Carus Monographs: Duane Bailey, Chairman (1985-86); Pete Hayslett (1985-87); John Karlof (1982-85).

Subcommittee on Dolciani Mathematical Expositions: Joseph Malkevitch, Chairman (1981-87); Donald J. McCarthy (1983-86); Donald B. Small (1982-85).

Subcommittee on Lester R. Ford Awards: Richard S. Millman, Chairman (1985); Paul R. Halmos (1983-86), ex-officio; Robert Osserman (1984-86); Stephen E. Puckette (1985-87).

Subcommittee on MAA Notes: Alan C. Tucker, Chairman (1982-85); Charles A. Cable (1982-86); William F. Lucas (1982-85).

Subcommittee on MAA Studies in Mathematics: Alan C. Tucker, Chairman (1980-85); James P. Crawford (1984-86); William F. Lucas (1985-87).

Subcommittee on New Mathematical Library: William G. Chinn, Chairman (1982-85); Donald J. Albers (1985); Basil Gordon (1980-85); Anneli Lax, ex-officio; Max M. Schiffer (1979-85).

Subcommittee on George Polya Awards: Larry A. Curnutt, Chairman (1982-85); Donald J. Albers (1982-87); Patrick Boyle (1984-86); Warren Page (1985-88), ex-officio.

Committee on Articulation:

Billy E. Rhoades, Chairman (1982-85); Philip C. Curtis, Jr. (1982-85); Ann D. Holley (1982-85); Chancey O. Jones (1982-85); Joan P. Leitzel (1982-85); Vincent O'Connor (1982-85); Barbara Wickless (1982-85).

Committee on Minicourses:

John O. Riedl, Chairman (1982-87); Sol Garfunkel (1984-86); Frederick Hoffman (1982-87); Warren Page (1982-85); Doris Schattschneider (1984-85); David A. Smith (1984-86).

Committee on Secondary School Lectures:

Henry Gore, Chairman (1982-85); Jane M. Day (1984-86); Allen H. Holmes (1985-87); John M. Jobe (1981-86); Amy C. King (1982-85); Gary J. Sherman (1982-85); David L. Skoug (1982-85); Donald B. Small, Advisor (1979-85).

Committee on Sections:

David W. Ballew, Chairman (1981-87); Yousef Alavi (1984-86), ex-officio; Sylvan Burgstahler (1984-86); Barbara T. Faires (1982-85); Glen E. Johnston (1985); Norman F. Lindquist (1985-86); John D. Neff (1985-87); John M. Smith (1985-86); Alfred B. Willcox, ex-officio.

Public Information Panel: Yousef Alavi, Chairman (1984-86); James C. Bradford (1984-86); Peter Hilton (1984-85); James R. Leitzel (1984-87); Jean J. Pedersen (1984-85); Marcia P. Sward, ex-officio.

Committee on Special Funds of the Association:

Lida K. Barrett, Chairman (1981-86); Leonard Gillman (1980-87), ex-officio; John M. Jobe (1982-87); Eileen L. Poiani (1982-87); G.B. Price (1985-86); David P. Roselle (1985-88); Richard M. Schori (1982-87).

Committee on the Greater MAA Fund:

Leonard Gillman, ex officio, Chairman; Ronald Davis, ex officio; Lida Barrett, ex officio; Ivan Niven, ex officio.

Committee on the Award for Distinguished Service to Mathematics:

Peter J. Hilton, Chairman (1985); Henry L. Alder (1984-86); Anneli Lax (1985-87).

Committee on the Chauvenet Prize:

Peter J. Hilton, Chairman (1985); Theodore W. Gamelin (1985-87); Lawrence A. Zalcman (1984-86).

Committee on the MAA Book Prize:

Doris Schattschneider, Chairman (1985); J.A. Seebach (1984-86); Gary J. Sherman (1985-87).

Committee on the Putnam Prize Competition:

Leonard F. Klosinski, Director (1980-86); Bruce Reznick, Chairman (1985); Gerald L. Alexanderson (1983-86); Loren C. Larson, Advisor (1984-86); Richard P. Stanley (1984-86); Harold M. Stark (1985-87).

Committee on the Teaching of Undergraduate Mathematics (CTUM):

Alan H. Schoenfeld, Chairman (1982-85); Donald W. Bushaw (1980-85); Gloria F. Gilmer (1981-86); Bonnie Gold (1985-87); Christopher Nevison (1984-86); Jean J. Pedersen (1984-86); Peter Ross (1985-87); Martha J. Siegel (1981-86).

CTUM-CCIME Panel on Computers and Technology in Mathematics Instruction: Ronald H. Wenger, Chairman; Dorothy Derringer; Gloria Gilmer; Christopher Nevison; Gerald Porter; David I. Schneider; J.A. Seebach; Martha J. Siegel; David Smith.

CTUM Panel on the Evaluation of Teaching: Donald W. Bushaw, Chairman; Robert B. Ash; Susanna S. Epp; Robert H. McDowell; Jean J. Pedersen; Arnold K. Pizer; James W. Vick.

CTUM Panel on Teaching Assistants and Part-Time Instructors: Bettye Anne Case, Chairman; Felix E. Browder; Wendell Fleming; John Philip Huneke; David Kraines; Alan H. Schoenfeld, ex-officio.

Committee on Two-Year Colleges (CTYC):

Ann E. Watkins, Chairman (1985-86); Donald J. Albers (1984-86); James J. Baldwin (1984-86); Bettye Anne Case (1985-88); Wade Ellis, Jr.

Ad Hoc Committee to Review the Visiting Lecturer Program:

John D. Neff, Chairman (1985-86); Lida K. Barrett (1985-86); John Bradburn (1985-86); Theodore S. Chihara (1985-86); William F. Lucas (1985-86).

Committee on Election Procedures:

Richard D. Anderson, Chairman (1984-86); Henry L. Alder (1984-86); Henry O. Pollak (1984-86).

Ad Hoc Committee on Summer Meetings:

Kenneth A. Ross, Chairman; Donald J. Albers; Lynn A. Steen; Marcia Sward; Alfred B. Willcox.

MAA Program Committee for Laramie Meeting:

Kenneth I. Gross, Chairman; David W. Ballew; Robert E. Gaines; A. Duane Porter.

Joint AMS-MAA Committee for Laramie Meeting:

George E. Andrews, Chairman (MAA); Jeanne L. Agnew (MAA); Paul F. Baum (AMS); William P. Ziemer (AMS).

MAA Program Committee for New Orleans Meeting:

Jerome A. Goldstein, Chairman; Garrett J. Etgen; O. Carruth McGehee; John L. Tilley.

Joint AMS-MAA Committee for New Orleans Meeting:

William C. Waterhouse, Chairman (AMS); George E. Andrews (MAA); Judith V. Grabiner (MAA); W. Gilbert Strang (AMS).

**Joint Committees**

Joint AMS-MAA-SIAM Committee on Employment Opportunities: (Terms expire October 31)

Patrick Hagan, Chairman (1985, SIAM); Wilfred E. Barnes (1985-87, MAA); Richard E. Ewing (1985-86, SIAM); Calvin T. Long (1985-86, MAA); John W. Petro (1985, AMS); Donald C. Rung (1985-87, AMS).

Joint MAA-ACM Committee on Retraining for Computer Science:

Donald L. Kreider, Chairman; Susan J. Devlin (MAA); Patrick Goldberg (ACM); Zaven Karian (MAA); Anthony Ralston (ACM); Olaf Stakelberg (ACM).

Joint Committee on Women in The Mathematical Sciences:

Carole B. LaCampagne, Chairman (1985-86); Nancy Angle (1985, NCTM); Raymond G. Ayoub (1985-86, AMS); Susan J. Devlin (ASA); Marjorie G. Hahn (IMS); Gloria C. Hewitt (1985-86; MAA); Julie Knight (1985-86, AMS); Betty K. Lichtenberg (1985-87, NCTM); Dianne P. O'Leary (1985, SIAM); Ingram Olkin (IMS); Katherine L. Pedersen (1985, NCTM); Linda Petzold (1985-86, SIAM); Judith Roitman (1985-86, AMS); Alice T. Schafer (1985-86, MAA); Elisabeth L. Scott (IMS).

AMS-MAA Joint Meetings Committee:

William J. LeVeque, AMS; Everett Pitcher, AMS; (1985-88); Ann D. Holley (1982-87); Charles David Miller (1985-87); Warren Page (1982-86), ex-officio; Stephen B. Rodi (1985-86); Jean B. Smith (1982-85); Larry Trivieri (1985-87).

Committee on the Undergraduate Program in Mathematics (CUPM):

Jerome A. Goldstein, Chairman (1983-87); Donald J. Albers (1984-86); Gerald L. Alexanderson (1980-85); Donald W. Bushaw (1982-85); Bettye Anne Case (1982-87); Philip C. Curtis, Jr. (1984-86); Kenneth I. Gross (1985-87); Stephen B. Maurer (1985-86); Mary A. Norton (1982-85); Billy E. Rhoades (1981-86); Alan H. Schoenfeld (1982-85); Lynn A. Steen (1985-86), ex-officio; W. Gilbert Strang (1985-87).

Joint CUPM-CCIME Panel on the Curricular Aspects of Computers in the Classroom: Carl Leinbach, Chairman; Sheldon P. Gordon; John M. Hosack; F. David Lesley; Stephen B. Maurer; Gerald Porter; J.A. Seebach.

CUPM Panel on the Continuing Mathematical Education of Teachers (COMET): Calvin T. Long, Co-Chairman; Bruce E. Meserve, Co-Chairman; Henry L. Alder; John A. Dossey; Marjorie M. Enneking; Richard K. Guy; Katherine P. Layton; Mary Ann Roberts Norton; Alan R. Osborne.

Joint CUPM-CTUM Panel on Service Courses: Donald W. Bushaw, Chairman; David W. Ballew; Jerome A. Goldstein; Martha J. Siegel; Maynard D. Thompson; James W. Vick.

Joint CUPM-CTUM-AMATYC Panel on the Curriculum at Two Year Colleges: Ronald Davis, Chairman; Geoffrey Akst; Donald J. Albers; Richard D. Anderson; John Bradburn; Bettye Anne Case; Larry Curnutt; Dale Ewen; Sheldon Gordon; Karl Smith.

Joint CUPM-CTUM-CTYC Panel on Remediation: Gloria F. Gilmer, Chairman; Donald J. Albers; Gerald L. Alexanderson; Carole Bauer; William Drezdzon; Eleanor Green Jones; John G. Harvey; Miriam R. Hecht; Peter Hilton; Anneli Lax; Jean J. Pedersen.

CUPM Panel on Calculus Articulation: Donald Small, Chairman; Gordon Bushaw; John H. Hodges; Donald Nutter; Ronald Schnackenberg; Donald Sherbert; Barbara Stott.

Joint MAA-ACM-ASEE Panel on Discrete Mathematics in the First Two Years: Martha J. Siegel, Chairman; Donald Bushaw; Jerome Goldstein; Anthony Ralston; Alfs T. Berztiss (ACM); Gerald L. Isaacs (ACM); John Schmeelk (ASEE); Stephen B. Maurer.

Committee on Computers in Mathematics Education (CCIME):

Gerald J. Porter, Chairman (1985); Susan Devlin (1985-86); Donald L. Kreider (1985-86); L. Carl Leinbach (1985-86); F. David Lesley (1985); Alan Schoenfeld (1985-87); J.A. Seebach (1985-87); Ronald Wenger (1985).

Committee on Visiting Lecturers and Consultants:

Theodore S. Chihara, Chairman (1981-86); Vivian Y. Kraines (1981-86); Jon M. Laible (1980-85); Eugene R. Seelbach (1979-87).

Kenneth A. Ross (1984-89), MAA; Alfred B. Willcox, MAA.

CEEB-MAA Joint Committee on Mutual Concerns:

John W. Kenelly, Co-Chairman (1985, CEEB); Kenneth A. Ross, Co-Chairman (1984-89, MAA); Henry L. Alder (1985, MAA); Chancey O. Jones (CEEB); Donald L. Kreider (CEEB); Lynn A. Steen (1985-86, MAA), ex-officio.

Joint Policy Board for Mathematics (JPBM):

I. Edward Block (SIAM), ex-officio; Hirsh G. Cohen (SIAM), ex-officio; Irving Kaplansky (AMS), ex-officio; William J. LeVeque (AMS), ex-officio; Kenneth A. Ross (MAA), ex-officio; Lynn A. Steen (MAA), ex-officio; Alfred B. Willcox (MAA), ex-officio.

JPBM Committee on the Status of the Profession:

Robert McKelvey, Chairman (SIAM); Lida K. Barrett (MAA); Wendell H. Fleming (SIAM); Murray Gerstenhaber (AMS); Louise Hay (AMS); Irwin Kra (member-at-large); Bernard L. Madison (MAA).

JPBM Committee for Mathematics Department Heads:

David Ballew, Chairman (1985-86); Donald Albers (1985); John Feroe (1985-87); John D. Fulton (1985-87); Paul D. Humke (1985-86); David J. Lutzer (1985-87); Bernard Madison (1985); Stephen B. Rodi (1985-86); Richard Schori (1985).

Joint AMS-MAA-SIAM Administrative Committee:

Everett Pitcher, Chairman (AMS); I. Edward Block (SIAM); Paul W. Davis (SIAM); Leonard Gillman (MAA); William J. LeVeque (AMS); Franklin P. Peterson (AMS); Kenneth A. Ross (MAA); Alfred B. Willcox (MAA).

JPBM Committee on Public Understanding of Mathematics:

Ronald L. Graham (AMS); Joseph B. Keller (SIAM); Lynn A. Steen (MAA).

Joint AMS-MAA-NAM Committee on Graduate Programs at Traditionally Black Institutions:

C.B. Bell (AMS); Lipman Bers (AMS); R. Creighton Buck (MAA); I.N. Herstein (MAA); Frank A. James (NAM); Ted Sykes (NAM).

**Editorial Boards of the Association**

American Mathematical Monthly (All terms expire December 31, 1986):

Editor: Paul R. Halmos.

Associate Editors: Gerald L. Alexanderson, Sabra S. Anderson, Sheldon Axler, Thomas F. Banchoff, Ralph P. Boas, David Borwein, W.W. Comfort, Hugh M.W. Edgar, Allan L. Edmonds, John H. Ewing, Richard K. Guy, Raoul Hailpern, Richard S. Millman, Dale H. Mugler, Seymour Schuster, J.A. Seebach, Lynn A. Steen, Mary R. Wardrop, Robert F. Wardrop, Herbert S. Wilf.

Mathematics Magazine (All terms expire December 31, 1985):

Editor: Doris Schattschneider.

Associate Editors: Edward J. Barbeau, John A. Beidler, Paul J. Campbell, Underwood Dudley, Gerald A. Edgar, Joseph A. Gallian, Judith V. Grabiner, Raoul Hailpern, Joseph Malkevitch, Pierre J. Malraison, William A. McWorter, Jr., Leroy F. Meyers, Jean J. Pedersen, Gordon Raisbeck, Ian Richards, Eric S. Rosenthal, Dan Shapiro, David A. Smith.

The College Mathematics Journal (All terms expire December 31, 1988):

Editor: Warren Page.

Associate Editors: Jeanne Agnew, Gerald L. Alexanderson, Patrick J. Boyle, Donald W. Bushaw, Michael W. Chamberlain, Leo Chosid, Josephine Edwards, Gino T. Fala, Stanley Friedlander, Robert M. Geist, Samuel A. Greenspan, Erwin Just, James J. Kaput, Alan Mackay, Peter L. Renz, Billie A. Rice, Ann E. Watkins, David H. Wheeler, Milton E. Winger.

Newsletter Editorial Committee:

Ronald M. Davis, Chairman (1981-85); Yousef Alavi (1981-86), ex-officio; Donald J. Albers, Associate Editor (1985); Philip M. Cheifetz (1981-86); William Chinn, Associate Editor (1981-85); Joseph A. Gallian (1981-86); Katherine P. Layton (1981-86); Stephen B. Maurer, Associate Editor (1983-86); John D. Neff (1985-87); Eileen L. Poiani (1981-85); Kenneth A. Ross (1984-89), ex-officio; Marcia Sward, ex-officio; Alan C. Tucker (1985); Paul Zorn, (1985-87).

**Representatives of the Association**

To Sections of AAAS: Section A: A.B. Willcox (1984-86); Section Q: A.B. Willcox (1984-86); Section U: A.B. Willcox (1984-86); Section X: A.B. Willcox (1984-86).

On the UMAP Editorial Board: Alphonse Buccino (1985-87); Courtney S. Coleman (1985-87); Peter A. Lindstrom (1985-87); Michael Olinick (1985-87).

On the COMAP Consortium Council: Alphonse Buccino (1985-87).

To the Conference Board of the Mathematical Sciences (CBMS): Lynn A. Steen (1985-86).

To the Council of Scientific Society Presidents: Lynn A. Steen (1985-86), ex-officio.

On the Discipline Committee for Mathematics of CEEB: Donald L. Kreider.

On the Governing Council of Mu Alpha Theta: Katherine P. Layton (1985-87).

MAA Representative on AMS Science Policy Committee: Lynn A. Steen (1985-86).

On the U.S. Commission on Mathematical Instruction (USCMI): Warren Page (1984-85); Eileen L. Poiani (1981-86).

On the April 1986 NCTM Program Committee: Ronald M. Davis.

7. A. Tenenbaum and M. Augenstein, Data Structures Using PASCAL, Prentice-Hall, Englewood Cliffs, NJ, 1981, pp. 149–154.

8. T. R. Walsh, The Towers of Hanoi Revisited: Moving the Rings by Counting the Moves, Information Processing Letters 15, 1982, pp. 64–67.

9. D. Wood, The Towers of Brahma and Hanoi Revisited, Computer Science Technical Report No. 80-CS-23, McMaster University, 1980.

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## A SHORT PROOF OF STEINHAUS' THEOREM ON SUMMABILITY

JEFF CONNOR

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242*

In this note, we offer a soft (and short!) proof of a result of Steinhaus: a regular matrix summability method cannot sum all sequences of 0's and 1's.

Let $x$ and $y$ be sequences. By a *matrix summability method* we mean a sequence to sequence transformation defined by an infinite matrix $A = (a_{n,k})_{n,k=1}^{\infty}$, an array of complex numbers, and $y = Ax$ when

$$y_n = \sum_{k=1}^{\infty} a_{n,k} x_k \quad \text{for all } n \in N.$$

$A$ is called regular if it maps convergent sequences into convergent sequences and preserves limits, i.e., $\lim_n (Ax)_n = \lim_n x_n$ if $x$ is convergent. Regular matrices are characterized by the following:

THEOREM (Silverman, Toeplitz) [**PS**]. *Let $A = (a_{n,k})_{n,k=1}^{\infty}$ be a matrix summability method. $A$ is regular if and only if*

$$(1) \quad \lim_n \sum_{k=1}^{\infty} a_{n,k} = 1,$$

$$(2) \quad \lim_n a_{n,k} = 0 \quad \textit{for all } k \in N, \textit{ and}$$

$(3)$ *there is an $M > 0$ such that*

$$\sum_{k=1}^{\infty} |a_{n,k}| < M \quad \textit{for all } n \in N.$$

*$A$ sums $x$* or *$x$ is $A$-summable* if $A$ maps $x$ into a convergent sequence.

We also use the Baire Classification Theorem, a direct ancestor of the Baire Category Theorem:

THEOREM (Baire) [**O**, p. 32]. *Let $X$ be a complete metric space and $(f_n)_{n=1}^{\infty}$ a sequence of continuous complex-valued functions. If $f(x) = \lim_n f_n(x)$ exists for all $x \in X$, then $f$ is continuous except at a set of points of first category.*

Let $\Omega$ denote the sequences of 0's and 1's and define a metric $d$ on $\Omega$ by

$$d(x,y) = \sum_{n=1}^{\infty} |x_n - y_n| \cdot 2^{-n}.$$

Observe that $(\Omega, d)$ is a complete metric space homeomorphic to the Cantor set, the metric topology being the topology of coordinatewise convergence in $\Omega$. We are now ready to give the promised proof.

THEOREM (Steinhaus) **[PS]**, **[S]**. *A regular matrix summability method cannot sum all sequences of* 0'*s and* 1'*s*.

*Proof*. Let $A = (a_{n,k})$ be regular. By (3), $\sum_{k=1}^{\infty}|a_{j,k}|$ is convergent for each $j \in N$, hence $f_j(x) = \sum_{k=1}^{\infty} a_{j,k} x_k$ is defined for all $x = (x_k) \in \Omega$, and $f_j$ is a continuous complex-valued function on $\Omega$ because $|f_j(x) - f_j(y)| \leqslant \sum_{k=i+1}^{\infty}|a_{j,k}| < \varepsilon$ whenever $d(x, y) < 2^{-i}$ and $i$ is sufficiently large.

Now, suppose $A$ sums all sequences of 0's and 1's, or in other words, $f(x) = \lim_j f_j(x)$ is defined for all $x \in \Omega$. Notice that any open ball about any point in $\Omega$ contains two points $z$ and $w$ where $z$ is eventually all 1's and $w$ is eventually all 0's. Since $A$ is regular, we have that $|f(z) - f(w)| = 1$. Hence, $f$ is discontinuous at every point of $\Omega$.

$f$, however, is the pointwise limit of a sequence of continuous functions, and consequently must be continuous at at least one point of $\Omega$. This is a contradiction. ∎

**References**

**[PS]** R. E. Powell and S. M. Shah, Summability Theory and Its Applications, Van Nostrand Reinhold, London, 1972.

**[O]** J. C. Oxtoby, Measure and Category, Springer-Verlag, New York, 1971.

**[S]** H. Steinhaus, Some remarks on the generalization of limit, Prace Mat. Fiz., 22 (1911) 121–134.

## A COMBINATORIAL CONSTRUCTION OF A NONMEASURABLE SET

ROBIN THOMAS

*Department of Mathematical Analysis, Charles University, Sokolovská 83, 186 00 Prague, Czechoslovakia*

We give a slightly unusual construction of a Lebesgue nonmeasurable set on the real line using a fundamental graph-theoretical assertion (Proposition 1).

First recall some graph-theoretical notions. A graph is a pair $(V, E)$, where $V$, the set of vertices, is an arbitrary set (possibly infinite) and $E$, the set of edges, is contained in $\binom{V}{2}$ (= the collection of all two-element subsets of $V$). A path in a graph $(V, E)$ is a sequence $v_0, \dots, v_n$ of distinct elements of $V$ such that $\{v_i, v_{i+1}\} \in E$ for $i = 0, 1, \dots, n-1$. The number of edges, $n$, is the length of the path. A cycle is a closed path, that is, $v_0 = v_n$. A graph $(V, E)$ is called bipartite if there are disjoint sets $V_1$ and $V_2$ such that $V_1 \cup V_2 = V$ and $\binom{V_i}{2}$ are disjoint from $E$, that is, neither $V_1$ nor $V_2$ contains both endpoints of an edge.

The following two propositions are well known, but for the reader's convenience their proofs are given here.

PROPOSITION 1. *Any graph* $(V, E)$, *which has no cycles of odd length, is bipartite.*

*Proof*. It suffices to prove the proposition for connected graphs only. Choose any $x \in V$ and define

$$V_1 = \{y \in V\colon \text{any path connecting } x \text{ and } y \text{ is of odd length}\},$$

$$V_2 = \{y \in V\colon \text{any path connecting } x \text{ and } y \text{ is of even length}\}.$$

Since there are no odd cycles in $(V, E)$, any two paths connecting $x$ and $y$ are of the same parity. Thus $V_1$ and $V_2$ yield the desired partition of $V$.

PROPOSITION 2. *Let $M$ be a Lebesgue measurable set of real numbers with positive measure. Then there exists a $\delta > 0$ such that $M \cap (x + M) \neq \varnothing$ for any $x \in \mathbb{R}$ with $|x| < \delta$.*

*Proof*. Find a closed set $F$ and an open set $G$ with $F \subseteq M$ and $F \subseteq G$ such that $3\lambda G < 4\lambda F$ (where $\lambda$ is the Lebesgue measure). Since $G$ is a countable union of disjoint open intervals, there is one among them, say $I$, such that $3\lambda I < 4\lambda(F \cap I)$. Let $\delta = \lambda I/2$ and suppose that $|x| < \delta$. Then $I \cup (x + I)$ is an interval of length less than $\frac{3}{2}\lambda I$ which contains both $F \cap I$ and $x + (F \cap I)$. The last two sets cannot be disjoint, since otherwise

$$\frac{3}{2}\lambda I = \frac{3}{4}\lambda I + \frac{3}{4}\lambda I < \lambda((F \cap I) \cup (x + (F \cap I))) \leqslant \lambda(I \cup (x + I)) < \frac{3}{2}\lambda I,$$

which is a contradiction. Hence,

$$\varnothing \neq (F \cap I) \cap (x + (F \cap I)) \subseteq M \cap (x + M),$$

which completes the proof.

We now proceed to the construction. Define a graph $(\mathbb{R}, E)$ as follows: $\{x, y\} \in E$ when $|x - y| = 3^k$ for some integer $k$. We claim that $(\mathbb{R}, E)$ contains no cycles of odd length. Indeed, suppose that $x_0, \ldots, x_n$ is a cycle of length $n$ in $(\mathbb{R}, E)$. Then for $i = 1, 2, \ldots, n$, $x_i = x_{i-1} + \xi_i 3^{k_i}$, where $\xi_i = \pm 1$ and each $k_i$ is an integer. Then

$$\xi_1 \cdot 3^{k_1} + \cdots + \xi_n \cdot 3^{k_n} = 0.$$

For sufficiently large $N$,

$$\xi_1 \cdot 3^{k_1 + N} + \cdots + \xi_n \cdot 3^{k_n + N}$$

is a sum of odd integers, so in order for the sum to be 0, there must be an even number of them. This shows that the cycle is even. By Proposition 1 there are disjoint sets $A$ and $B$ with $\mathbb{R} = A \cup B$ such that neither $A$ nor $B$ contains both endpoints of an edge. Since $x$ and $x + 3^k$ are joined by an edge, we have $A + 3^k \subseteq B$ and $B + 3^k \subseteq A$ for any integer $k$. Now by Proposition 2 the sets $A$ and $B$ are nonmeasurable.

REMARKS. (i) In fact we see that $A + 3^k = B$ and $A - 3^k = B$ for any integer $k$. This follows from the inclusions

$$A = A + 3^k - 3^k \subseteq B - 3^k \subseteq A,$$
$$B = B + 3^k - 3^k \subseteq A - 3^k \subseteq B.$$

(ii) The same idea has recently been used by Nešetřil and Rődl [**1**] to obtain a short proof of the existence of non-Ramsey sets.

EXERCISE. It is known that any construction of a nonmeasurable set must involve the Axiom of Choice. At which step of our construction is the Axiom of Choice necessary?

**Reference**

1. J. Nešetřil and V. Rődl, Two Remarks on Ramsey Theorem, to appear in Discrete Math.

## WHEN IS AN ORDINARY DIFFERENTIAL EQUATION SEPARABLE?

DAVID SCOTT

*Department of Mathematics and Computer Science, University of Puget Sound, Tacoma, WA 98416*

Elementary differential equations texts generally present the method of separation of variables by stating that if one can write $f(x, y) = g(x)/h(y)$, then the equation $y' = f(x, y)$ can be solved by integrating each side of $h(y)\,dy = g(x)\,dx$. They don't raise the question of whether $f(x, y)$ factors into a product of a function of $x$ times a function of $y$. Most of the exercises give $f(x, y)$ in factored form, and many students cannot solve $y' = 1 + x + y^2 + xy^2$ unless they are told it is separable.

Most texts mention that a separable equation gives rise to the exact differential $g(x)\,dx - h(y)\,dy$, while treating separation of variables and exact differentials as distinct methods. The

sections on exact equations contain the standard test for exactness and often include a discussion of integrating factors, but I have seen none that develops a test for separability.

This paper gives a simply stated test for separability. In what follows $D$, the domain of $f$, is an open convex set in the plane, $f$ is real valued, and partial derivatives of $f$ are indicated by subscripts.

PROPOSITION 1. *If there are differentiable functions $\phi(x)$ and $\psi(y)$ such that $f(x, y) = \phi(x)\psi(y)$ for all $(x, y) \in D$, then $f(x, y)f_{xy}(x, y) = f_x(x, y)f_y(x, y)$ for all $(x, y) \in D$.*

The proof follows directly from $f_x(x, y) = \phi'(x)\psi(y)$ and the corresponding formulas for the other partial derivatives.

PROPOSITION 2. *If $f$, $f_x$, $f_y$ and $f_{xy}$ are continuous in $D$, $f(x, y)$ is never $0$ in $D$, and $f(x, y)f_{xy}(x, y) = f_x(x, y)f_y(x, y)$ for all $(x, y) \in D$, then there are continuously differentiable functions $\phi(x)$ and $\psi(y)$ such that $f(x, y) = \phi(x)\psi(y)$ for all $(x, y) \in D$.*

*Proof.* $f$ has the same sign throughout $D$ and factoring $f$ is equivalent to factoring $-f$, so we may assume that $f$ is positive on $D$. Now

$$\frac{\partial}{\partial y}\frac{f_x(x, y)}{f(x, y)} = \frac{f(x, y)f_{xy}(x, y) - f_x(x, y)f_y(x, y)}{f^2(x, y)} = 0.$$

Since $\frac{\partial}{\partial x}\ln(f(x, y)) = f_x(x, y)/f(x, y)$, we conclude that $\frac{\partial}{\partial x}\ln(f(x, y))$ is a function of $x$ alone and can write $\frac{\partial}{\partial x}\ln(f(x, y)) = \alpha(x)$. The function $\alpha$ is continuous because $f$ and $f_x$ are.

Let $\beta(x) = \int\alpha(x)\,dx$. Then there is a function $\gamma$ such that $\ln(f(x, y)) = \beta(x) + \gamma(y)$. $\gamma$ is continuously differentiable because $f$ and $f_y$ are continuous.

Let $\phi(x) = \exp(\beta(x))$ and $\psi(y) = \exp(\gamma(y))$. Then $f(x, y) = \phi(x)\psi(y)$, completing the proof.

As the following example shows, the condition that $f$ is nonzero on $D$ cannot be dropped. Define $f(x, y)$ by

$$f(x, y) = \begin{cases} x^2\exp(y), & \text{if } x \geqslant 0, \\ x^2\exp(2y), & \text{if } x \leqslant 0. \end{cases}$$

Direct computation shows that on $R^2$ we have $ff_{xy} = f_x f_y$. However $f$ does not factor since if it did $f(1, y)$, which is $\exp(y)$, and $f(-1, y)$, which is $\exp(2y)$, would be proportional as functions of $y$.

One can replace the condition that $f$ is nonzero on $D$ by requiring that $f(x, y)$ be an analytic function of $x$ and $y$. Then, using Proposition 2 and the fact that if two analytic functions agree in a neighborhood of a point in $D$ they agree on the connected set $D$, one can show that $ff_{xy} = f_x f_y$ implies that $f$ factors.

## TOPOLOGICAL PARTITIONS OF EUCLIDEAN SPACE BY SPHERES

PAUL BANKSTON
*Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI 53233*

RALPH FOX
*Department of Mathematics, George Mason University, Fairfax, VA 22030*

A space $X$ *partitions* a space $Y$ if there is a family of topological embeddings of $X$ into $Y$ whose images form a cover of $Y$ by pairwise disjoint sets. In [**1**] (whose notation and terminology we follow here) it is shown that the $n$-sphere $S(n)$ cannot partition euclidean $(n + 1)$-space $\mathbb{R}^{n+1}$ (an easy result) and that if $X$ is any nonempty subspace of $S(n)$, then $X$ partitions $\mathbb{R}^{2n+1}$. An obvious question then arises as to whether $S(n)$ partitions $\mathbb{R}^m$ for $n + 1 < m < 2n + 1$ (Ques-

tion 3.1 (iv) of [**1**]). In this note we give an affirmative answer based on an unpublished construction due to S. Kakutani and communicated to us by P. R. Halmos [**4**] (who learned of the construction forty years ago). This construction, used to give a partition of $\mathbb{R}^3$ by $S(1)$, is extremely simple, essentially different from those given in [**1**], and easily generalizable. Because several people have looked at the problem since the completion of [**1**], we thought the solution should be publicized as a short postscript to that paper.

THEOREM. *$S(n)$ partitions $\mathbb{R}^{n+2}$ in such a way that each sphere is tamely embedded, unknotted, and not linked with any other member of the partition.*

*Proof.* We partition $\mathbb{R}^{n+2}$ into $n$-spheres as follows: First let $C_1$ be an open unbounded cylinder in $\mathbb{R}^{n+2}$. To be specific, we let $X_1$ be a coordinate axis in $\mathbb{R}^{n+1}$ and let $C_1$ be the product of $X_1$ with the open unit ball in the $(n+1)$-dimensional vector subspace perpendicular to $X_1$. Now $\mathbb{R}^{n+2} \setminus C_1$, being homeomorphic with $\mathbb{R}^1 \times [1, \infty) \times S(n)$, is easily partitioned into spheres; and $C_1$ is homeomorphic with $\mathbb{R}^{n+2}$.

Let $C_2$ be an open unbounded cylinder in $C_1$ formed in a $U$-shape—i.e., let $R_1, R_2$ be parallel translates of a closed ray in $X_1$, let $L$ be the straight line segment joining the endpoints of $R_1, R_2$, and let $C_2$ be the interior of a regular thickening of $R_1 \cup L \cup R_2$. We then choose a branch of $C_2$ and form another open unbounded $U$-shape $C_3$ in exactly the same fashion, making sure that the base of $C_3$ is at least a unit distance from the base of $C_2$. Proceed inductively. For each $n \geqslant 1$, $C_n$ is homeomorphic with $\mathbb{R}^{n+2}$, and $C_n \setminus C_{n+1}$ is homeomorphic with $\mathbb{R}^1 \times [1, \infty) \times S(n)$. Finally, because the $U$-shapes are pushed unboundedly far from the origin, it is easy to see that $\bigcap_{n=1}^{\infty} C_n = \varnothing$. This completes the construction. (It is obvious that there is no linking or knotting, and each sphere is tamely embedded.) □

The following question now becomes inevitable.

*Revised Question* 3.1 (iv) [**1**]: Is there a nonempty subspace of $S(2)$ which does not partition $\mathbb{R}^4$? Of course by the above we now need only look in $\mathbb{R}^2 \subseteq S(2)$ for candidates, but this brings us to the question of whether in general there are nonempty subsets of $\mathbb{R}^n$ which do not partition $\mathbb{R}^{n+1}$. This is basically Question 3.1 (ii) in [**1**], and we are little closer to a solution now than we were when the issue first arose.

REMARK. In [**5**], A. Szulkin offers a new solution to the problem of whether $S(1)$ partitions $\mathbb{R}^3$, in response to a re-posing of the question by H. S. Shapiro. (Problems like this have a way of surfacing time and again. See also [**2**] and [**3**] for similar problems.) So far, we know of four distinct solutions, each with its own particular merits. To comment:

(i) Of the two solutions offered in [**1**], one of them uses the Axiom of Choice. While the constructive solution is more visual, the Jordan curves in the partition are quite general. However, in the nonconstructive solution, all of the Jordan curves can be taken to be standard circles (of equal radius). Neither solution generalizes to higher dimensions.

(ii) Szulkin's solution combines the attractive features of both of the solutions in [**1**]; namely it constructively partitions $\mathbb{R}^3$ into circles (whose radii include all positive real numbers). However, it too fails to generalize to higher dimensions.

(iii) While the Kakutani solution (as well as the above generalization) is constructive, it involves highly noncircular Jordan curves. This, of course, prompts the question of whether $\mathbb{R}^{n+2}$ can always be partitioned into standard $n$-spheres.

**References**

1. P. Bankston and R. McGovern, Topological partitions, General Topology and Appl., 10 (1979) 215–229, MR # 80i:54007.
2. ______, On partitions of plane sets into simple closed curves, Proc. Amer. Math. Soc., 88 (1983) 691–697.
3. ______, On partitions of plane sets into simple closed curves, II, Proc. Amer. Math. Soc., 89 (1983) 498–502.
4. P. R. Halmos (private communication).
5. A. Szulkin, $\mathbb{R}^3$ is the union of disjoint circles, this MONTHLY, 90 (1983) 640–641.

THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

# A NON-SIMPSONIAN USE OF PARABOLAS IN NUMERICAL INTEGRATION

ARTHUR RICHERT

*Department of Mathematics, Southern College, Collegedale, TN 37315*

This paper presents a numerical integration technique which uses parabolas differently than does Simpson's rule. The degree of precision of the technique is three, as is the case for Simpson's rule, while the error bound is an improvement over Simpson's rule by a factor of almost eleven. The technique provides an example of the usefulness of Taylor polynomials in numerical analysis which may be presented to beginning calculus students as soon as Taylor's theorem has been covered.

In Simpson's rule, the interval of integration, $[a, b]$, is divided into an even number, $n$, of equal subintervals each of length $h = (b - a)/n$. The approximation is then given by

$$\int_a^b f(x)\,dx \approx \frac{h}{3}[f(x_0) + 4f(x_1) + 2f(x_2) + \cdots + 4f(x_{n-1}) + f(x_n)], \tag{1}$$

with an error bound of

$$|\text{Error}| \leqslant h^4 M_4 (b - a)/180, \tag{2}$$

where $M_4$ is an upper bound on the magnitude of $f^{(4)}$ over the interval $[a, b]$.

We begin in the same manner with the exception that $n$ is not required to be even. As an approximation to $f$ over the $i$th subinterval, $[x_{i-1}, x_i]$, we use the second degree Taylor polynomial for $f$ expanded about $X_i$, the midpoint of the $i$th subinterval. This polynomial is given by

$$T_i(x) = f(X_i) + f^{(1)}(X_i)(x - X_i) + f^{(2)}(X_i)(x - X_i)^2/2.$$

Direct integration yields

$$\int_{x_{i-1}}^{x_i} T_i(x)\,dx = h \cdot f(X_i) + \frac{h^3}{24} \cdot f^{(2)}(X_i). \tag{3}$$

The approximation analogous to (1) is obtained by summing (3) over the $n$ subintervals to get

$$\int_a^b f(x)\,dx \approx h \sum_{k=1}^{n} f(X_k) + \frac{h^3}{24} \sum_{k=1}^{n} f^{(2)}(X_k).$$

This formula is of interest in its lack of dependence on $f^{(1)}$ and also in that it is merely the midpoint rule plus a "correction" term involving the second derivative.

To determine an error bound we assume $f \in C^4[a, b]$ and express $f$ over the $i$th subinterval in the form

$$f(x) = \sum_{k=0}^{3} \frac{f^{(k)}(X_i)}{k!}(x - X_i)^k + \frac{f^{(4)}(c_x)}{24}(x - X_i)^4,$$

where for each $x \in [x_{i-1}, x_i]$ an appropriate $c_x$ between $x$ and $X_i$ is guaranteed by Taylor's theorem. Direct integration utilizing the weighted mean value theorem for integrals yields

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = h \cdot f(X_i) + \frac{h^3}{24} \cdot f^{(2)}(X_i) + \frac{h^5}{1920} \cdot f^{(4)}(d_i),$$

where $d_i \in (x_{i-1}, x_i)$. Thus the absolute error over the $i$th subinterval is bounded by $h^5M_4/1920$ and the error expression analogous to (2) is

$$|\text{Error}| \leqslant nh^5M_4/1920 = h^4M_4(b-a)/1920.$$

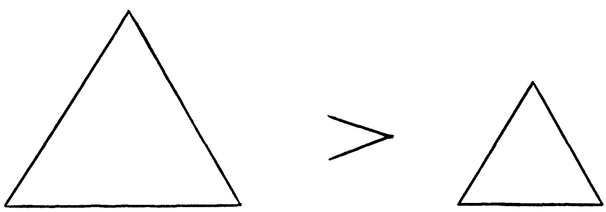
The degree of precision remains three while the error bound is reduced by a factor of nearly eleven. The cost of this improvement is the availability of $f^{(2)}$ and an additional $n-1$ evaluations.

---

## REBUSES AGAIN

According to the American Heritage Dictionary, a *rebus* is a riddle composed of words or syllables depicted by symbols or pictures that suggest the words or syllables they represent. Can you decipher the ones below?

1.



—Robert J. Beeber

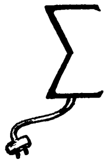
2. **XIЯTAM A**

—W. Edwin Clark

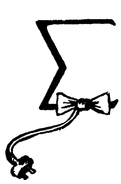
3. $3\int(\text{ice})^2\,d(\text{ice})$

—Wayne M. Dymacek

4a.

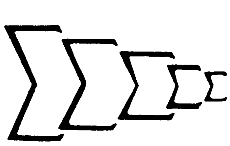


4b.

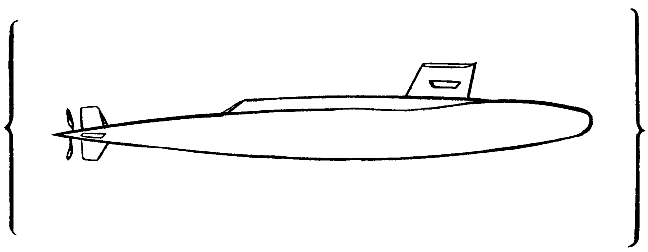


Dale T. Hoffman

5.



—Dale T. Hoffman

6.  —Daniel J. Sprows

Answers on p. 440.

---

## PROBLEMS AND SOLUTIONS



*Send all* ***proposed*** *problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions*** *should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

### ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by November 30, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3093. *Proposed by William C. Waterhouse, Pennsylvania State University.*

Let $f(x_1, \dots, x_n)$ be a real symmetric polynomial, and suppose that $f$ has an absolute minimum. Bunyakovskii showed long ago that the minimum need not occur at a point where all $x_i$ are equal (cf. this MONTHLY, 90 (1983) 378–387). Prove that $\nabla f = 0$ at some point where all $x_i$ are equal. Show on the other hand that this statement need not be true for an arbitrary $C^\infty$ symmetric function having an absolute minimum.

E 3094. *Proposed by Louis Funar, University of Craiova, Romania.*

Let $D$ be a convex body in $\mathbb{R}^3$, and let $\lambda(D) = \sup_\pi \text{area}\,(\pi \cap D)$ where $\pi$ is a variable plane. Prove that $D$ can be divided into four parts $D_1, D_2, D_3, D_4$ so that $\lambda(D_i) < \lambda(D)$, $i = 1, 2, 3, 4$. Is it possible to divide $D$ into three parts with the same property?

he obtains Presburger's decision procedure for arithmetic without multiplication and, in addition, he shows that the problem of deciding whether a given computational procedure will ever halt is algorithmically undecidable. Many details are left as exercises, and fairly detailed outline solutions are presented in an additional 50 pages. The book is intended for quite unsophisticated readers, is very clearly written, and should provide a very effective teaching tool.

In order to accomplish his task, Fisher required not only a logical formalism for elementary number theory, but also a model of computation for giving a precise definition of *computable* function. From the many mathematically equivalent formalisms available for the purpose, Fisher chose "register" machines. This simple and elegant formalism is highly accessible, especially to readers with a modicum of programming experience. However, because Fisher is working with a formalization of number theory in which $+$ and $\cdot$ are the only functions explicitly available, the task of proving that all computable functions are representable (in a suitable sense) in his formalization is rather formidable. Fisher uses the now classical device, Gödel's $\beta$ function:

$$\beta(c,d,i) = \mathrm{rm}(c, 1 + (i+1)d),$$

where $\mathrm{rm}(x, y)$ is the least nonnegative remainder when $x$ is divided by $y$. Using the Chinese remainder theorem, it is not too difficult to show that for every finite sequence $a_0, a_1, \dots, a_n$, there are integers $c, d$ such that $\beta(c,d,i) = a_i$ for $i = 0, 1, \dots, n$. The successive steps of a computation by a register machine program can then be coded using this $\beta$ function. The catch is that it is not much fun to formalize this in Peano arithmetic.

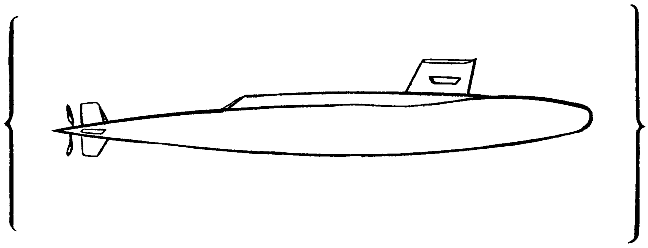
An alternative path, which avoids much of the tedious complications, is first to prove as a theorem of ordinary mathematics that computability by register machines is equivalent to a more tractable notion. For example, by a theorem of Kleene (later improved by Julia Robinson), easily proved by an *informal* use of Gödel's $\beta$ function, the computable functions are just those obtainable from a few initial functions using the two operations of function composition, and taking the least zero of a function. Another attractive possibility only recently available is to give the new Jones-Matijasevič proof of the Davis-Putnam-Robinson theorem that the graph of a computable function consists of the tuples which satisfy an exponential Diophantine equation. This new proof is very simple and is carried out using register machines. If this path is taken, the exponential function should probably be added to $+$ and $\cdot$ as primitive functions for which the recursion equations are taken as axioms. The slight weakening of the Gödel-Rosser theorem that would result hardly matters at this level, and as a dividend one would obtain the unsolvability of the decision problem for exponential Diophantine equations. (Alternatively, one could digress to obtain Matijasevič's theorem that the exponential function is Diophantine, thereby not only avoiding the weakening of the Gödel-Rosser theorem, but also obtaining the unsolvability of Hilbert's tenth problem.)

Two minor errata: On p. 34 it is stated that Russell and Whitehead's *Principia Mathematica* was written independently of Frege's work, which is certainly not the case. On p. 37 it is stated that there is no mechanical procedure for "deciding whether, for arbitrary $n$, a sequence of $n$ successive 7s occurs in the decimal expansion of $\pi, \dots$." What is presumably meant is that no such procedure is *known*.

---

## ANSWERS TO REBUSES ON PAGES 426–427

1. Triangle inequality.
2. The inverse of a matrix.
3. Iceberg (= ice cube + sea = $(\text{ice})^3 + C$).
4a. Power series.
4b. Formal power series.
5. Telescoping series.
6. Subset.

6.  —Daniel J. Sprows

Answers on p. 440.

## PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

### ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by November 30, 1985. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3093. *Proposed by William C. Waterhouse, Pennsylvania State University.*

Let $f(x_1, \ldots, x_n)$ be a real symmetric polynomial, and suppose that $f$ has an absolute minimum. Bunyakovskii showed long ago that the minimum need not occur at a point where all $x_i$ are equal (cf. this MONTHLY, 90 (1983) 378–387). Prove that $\nabla f = 0$ at some point where all $x_i$ are equal. Show on the other hand that this statement need not be true for an arbitrary $C^\infty$ symmetric function having an absolute minimum.

E 3094. *Proposed by Louis Funar, University of Craiova, Romania.*

Let $D$ be a convex body in $\mathbb{R}^3$, and let $\lambda(D) = \sup_\pi \text{area}\,(\pi \cap D)$ where $\pi$ is a variable plane. Prove that $D$ can be divided into four parts $D_1, D_2, D_3, D_4$ so that $\lambda(D_i) < \lambda(D)$, $i = 1,2,3,4$. Is it possible to divide $D$ into three parts with the same property?

E 3095. *Proposed by Alexandru Lupaş, Facultatea de Mecanică, Sibiu, Romania.*

Let $p$ be a fixed natural number, $(a_n)_{n=0}^{\infty}$ a sequence of real numbers, and

$$\Delta^p a_n = \sum_{k=0}^{p} (-1)^{p-k}\binom{p}{k} a_{n+k}, A_n = \frac{1}{n+1}\sum_{k=0}^{n} a_k.$$

If there exist constants $m_p, M_p$ such that $m_p \leqslant \Delta^p a_n \leqslant M_p$, $n = 0,1,2,\ldots$, prove that

$$\frac{1}{p+1} m_p \leqslant \Delta^p A_n \leqslant \frac{1}{p+1} M_p, \quad n = 0,1,\ldots.$$

E 3096. *Proposed by the editors* (cf. E2903 [1980, 619; 1983, 483; this issue]).

Consider paths in the northeast quadrant which start at $(0,0)$, such that each step is either one unit east or one unit north, and end at $(2n,2n)$. A path is called a Whitworth path if it does not go above the diagonal $y = x$. A path is called a Shapiro path if it avoids the points $(1,1),(3,3),\ldots,(2n-1,2n-1)$, i.e., those on the diagonal with odd coordinates. Exhibit, for fixed $n$, a one-to-one correspondence between the set of Whitworth paths and the set of Shapiro paths.

E 3097. *Proposed by Florentin Smarandache, Lycée Sidi El Hassan Lyoussi, Sefrou, Morocco.*

Find all real numbers $x$ for which

$$(x+1)^x + (x+2)^x = (x+3)^x.$$

E 3098. *Proposed by Roger Cuculière, Paris, France.*

Given two circles with diameters $IA = a$ and $IB = b$, and a set of smaller circles between them as in the following figures, find the total area enclosed by the small shaded circles in each of the following cases:

(a) The center of one of the small circles lies on the (common) diameter of the large circles (Fig. 1).

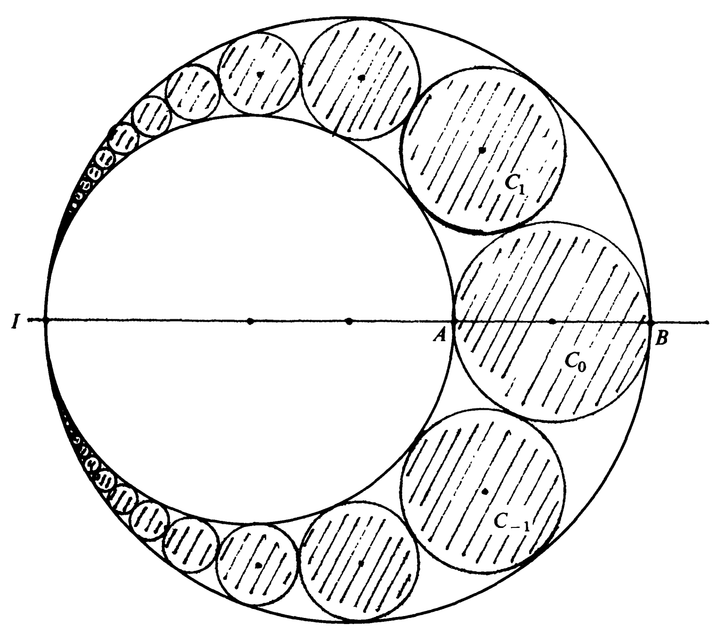



FIG. 1

(b) Two of the small circles are tangent to the diameter of the large circles (Fig. 2).

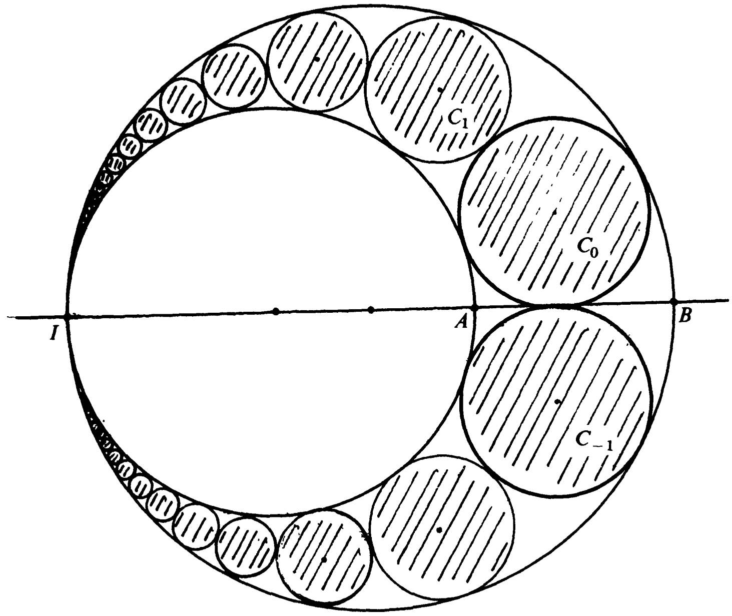



FIG. 2

(c) The case with no restrictions (Fig. 3).

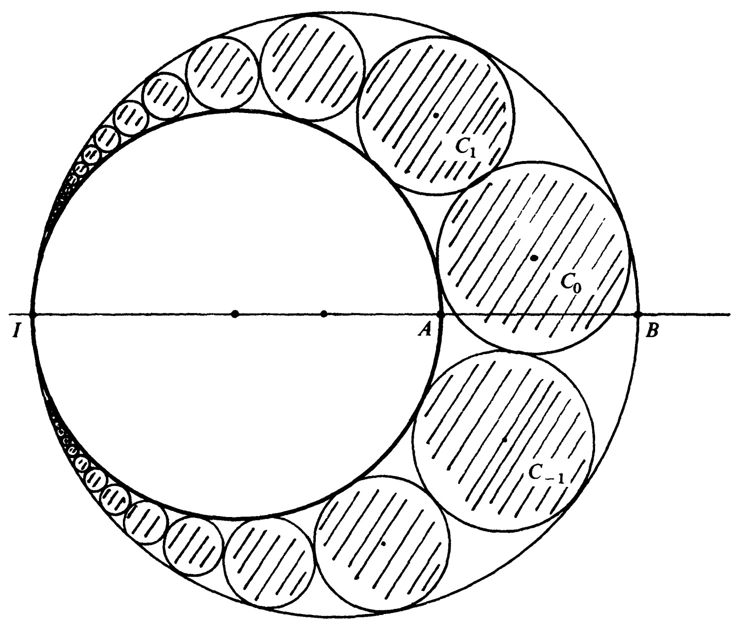



FIG. 3

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Random Walks and Catalan Numbers

E 2903 [1981, 619; 1983, 483]. *Proposed by Louis W. Shapiro, Howard University.*

Consider walks in the northeast quadrant which start at $(0,0)$ and such that each step is one unit east or north and the points $(1,1),(3,3),\ldots,(2k+1,2k+1),\ldots$ are forbidden. How many paths are there to $(2n,2n)$?

*Solution by D. M. Bloom, Brooklyn College, and the proposer.* By a "good" path we mean one in which each step is one unit north or east. For $n \geqslant 0$, let $c_n$ be the number of good paths from $(0,0)$ to $(n,n)$ which never go above the diagonal $x=y$; for $n \geqslant 1$, let $d_n$ be the number of good paths from $(0,0)$ to $(n,n)$ which lie entirely below that diagonal (except of course at the endpoints); and let $b_{2n}$ be the number which the problem asks for. By considering the least $k>0$ for which a given path meets the point $(k,k)$, we easily see that

$$c_n = \sum_{k=1}^{n} d_k c_{n-k} \qquad (\text{with } c_0 = 1),$$

$$b_{2n} = 2\sum_{k=1}^{n} d_{2k} b_{2n-2k} \qquad (\text{with } b_0 = 1).$$

Since evidently $d_k = c_{k-1}$, the above becomes

$$(1) \qquad c_n = \sum_{j=0}^{n-1} c_j c_{n-1-j}; \quad b_{2n} = 2\sum_{j=0}^{n-1} c_{2j+1} b_{2(n-1-j)}.$$

A straightforward induction using (1) shows that $b_{2n} = c_{2n}$ for all $n \geqslant 0$. But $c_n$ is the $n$th Catalan number $\binom{2n}{n}/(n+1)$.* Hence

$$b_{2n} = \binom{4n}{2n}\bigg/(2n+1).$$

(*The solution to E 2253 (this MONTHLY, 1971, p. 797) gives Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, Second Edition, p. 71, as a reference for this result. The result is also obtainable as a special case of the ballot problem.)

*Comment by the editors*: This solution is published in response to a letter by Professor Ira Gessel of MIT who writes

> "... The published solution to AMM problem E 2903, 'Random Walks and Catalan Numbers', which appeared in the August–September 1983 MONTHLY ... seems to claim that the number of paths from $(0,0)$ to $(n,n)$ which avoid certain points depends only on the number of 'options excluded.' This is clearly false. For example, the number of paths from $(0,0)$ to $(n,n)$ which avoid the point $(a,b)$ is $\binom{2n}{n} - \binom{a+b}{a}\binom{2n-a-b}{n-a}$, which is not independent of $a$ and $b$. However, if $0<a$, $b<n$ there are always exactly two 'options excluded'."

We ask in Problem E 3096, this issue, for problem solvers to reconsider this problem.

### Two Similar Triangles with a Common Circumcenter

E 2974 (corrected) [1982, 756; 1983, 482]. *Proposed by Jordi Dou, Barcelona, Spain.*

This problem, posed originally in the December 1982 issue of this MONTHLY, page 756, contained an error introduced in the editorial process. The problem should read: "Let $AMB$

(oriented clockwise) and $CMD$ (counter-clockwise) be similar triangles. Prove that triangles $ACX$ (clockwise) and $YDB$ (clockwise), both similar to the first triangles, have the same circumcenter."

*Solution* I *by Howard Eves, University of Maine.* We shall employ the analytic geometry of the complex plane, and follow the custom of designating points and their affixes by corresponding upper and lower case letters respectively. For convenience we choose our coordinate system so that the circumcircle of triangle $ACX$ is the unit circle centered at the origin. We then have $a\bar{a} = c\bar{c} = x\bar{x} = 1$. Since triangle $ACX$ is directly similar to triangle $AMB$, inversely similar to triangle $CMD$, and directly similar to triangle $YDB$, we have, in turn,

$$\begin{vmatrix} a & a & 1 \\ m & c & 1 \\ b & x & 1 \end{vmatrix} = 0, \begin{vmatrix} c & \bar{a} & 1 \\ m & \bar{c} & 1 \\ d & \bar{x} & 1 \end{vmatrix} = 0, \begin{vmatrix} y & a & 1 \\ d & c & 1 \\ b & x & 1 \end{vmatrix} = 0.$$

From the first equation we obtain

$$m = \frac{b(c-a) + a(x-c)}{x-a}.$$

From the second equation, after eliminating $m$ and simplifying, we obtain

$$d = b\frac{(c-a)(\bar{x}-\bar{a})}{(\bar{c}-\bar{a})(x-a)}.$$

From the third equation, after eliminating $d$ and simplifying, we obtain

$$y = b\frac{(c-a)(\bar{x}-\bar{c})}{(\bar{c}-\bar{a})(x-c)}.$$

We note that $y\bar{y} = d\bar{d} = b\bar{b}$, and the origin is the circumcenter of triangle $YDB$. Triangles $ACX$ and $YDB$ thus share the origin as a common circumcenter.

In the particular case where triangle $AMB$ is right-angled at $M$, it follows that the line segments $AX$ and $YB$ bisect each other.

*Solution* II *by John Oman, University of Wisconsin-Oshkosh.* A straightforward proof can be given using complex numbers and the following formula for the circumcenter $Z_c$ of the triangle whose vertices are the complex numbers $Z_1$, $Z_2$, and $Z_3$:

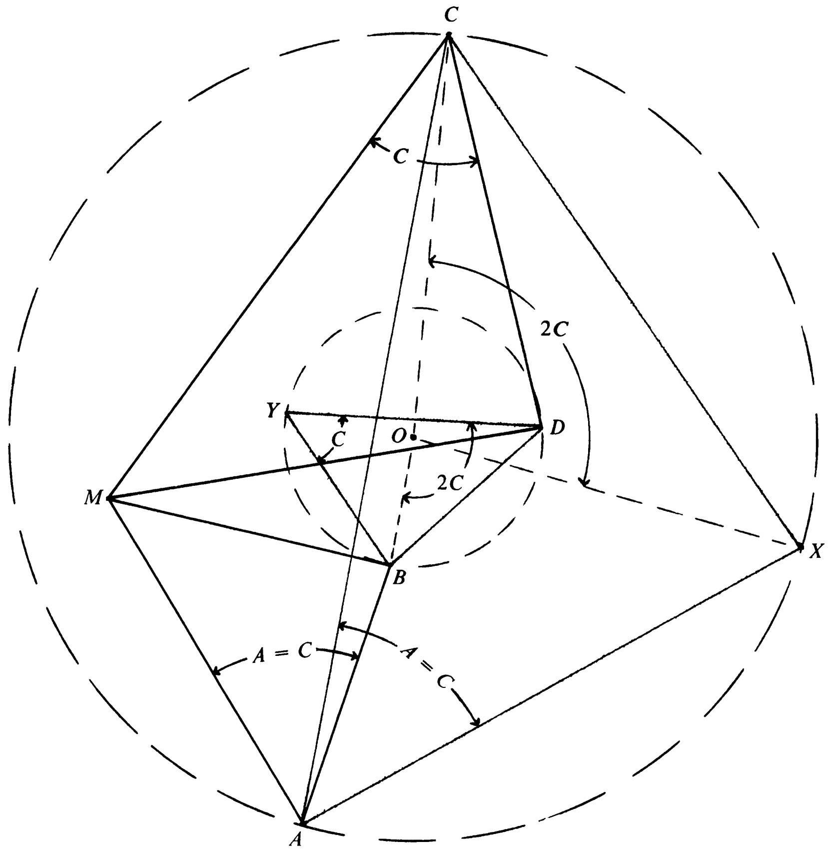
$$Z_c = \frac{\begin{vmatrix} Z_1\bar{Z}_1 & Z_2\bar{Z}_2 & Z_3\bar{Z}_3 \\ Z_1 & Z_2 & Z_3 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} \bar{Z}_1 & \bar{Z}_2 & \bar{Z}_3 \\ Z_1 & Z_2 & Z_3 \\ 1 & 1 & 1 \end{vmatrix}}$$

If $M, A, B, C$ are represented by the complex numbers 0, 1, $\Gamma$, and $\beta$, respectively, then $D, X, Y$ are given by $\beta\bar{\Gamma}$, $(1-\beta)\Gamma + \beta$, and $(\Gamma - \beta\bar{\Gamma})\frac{1}{\bar{\Gamma}} + \beta\bar{\Gamma}$. The circumcenter for both triangle $ACX$ and triangle $YDB$ is

$$\frac{\Gamma - \beta\bar{\Gamma} + \beta\Gamma\bar{\Gamma} - \Gamma\bar{\Gamma}}{\Gamma - \bar{\Gamma}}.$$

*Solution* III *by the proposer.* Let $S_A$ be the similarity of center $A$, angle $\sphericalangle BAM$, and ratio $AM/AB$, and $S_C$ the similarity of center $C$, angle $\sphericalangle MCD = \sphericalangle A$, and ratio $CD/CM$ $(= AB/AM)$. We then have that $R = S_A \cdot S_C$ is a rotation of angle $2A$, $R(B) = D$, and $R(X) = C$ since $S_A(X) = C$. If $O$ is the center of $R$, we have $\sphericalangle BOD = \sphericalangle XOC = 2A$. Because $\sphericalangle BYD = \sphericalangle XAC$

$= A$, they are inscribed angles in the circles having $\sphericalangle BOD$ and $\sphericalangle XOC$ as central angles, and hence $O$ is the circumcenter of the two triangles $YDB$ and $ACX$.




Also solved by A. Bondesen (Denmark), W. Janous (Austria), O. P. Lossers (The Netherlands), J. -P. Monier (France), H. J. Morse, and W. A. Newcomb.

**A Hereditary Property of Cyclic Linear Mappings**

E 2978 [1982, 757]. *Proposed by U. Abel, DKFZ, Heidelberg, Germany.*

A mapping $g$ is called cyclic if $P(g) = \min\{k | g^k = \text{identity}\}$ exists ($g^k = g \circ \cdots \circ g$, $k$ times).

Let $f$: $\mathbb{R}^n \to \mathbb{R}^n$ be a cyclic linear mapping, $R^n = V \oplus W, V, W$ being subspaces of $\mathbb{R}^n$, and $f(V) = V$.

Define $\tilde{f} = \text{Pr}_W \circ f|_W$: $W \to W$ ($\text{Pr}_W$ denoting the projection on $W$ along $V$). Then $\tilde{f}$ is cyclic, and $P(f) = \text{lcm}(P(\tilde{f}), P(f|_V))$.

*Solution by N. Miku, Catholic University, The Netherlands.* On a suitable basis the matrix $A$ of $f$ has the form $\begin{pmatrix} P & Q \\ 0 & R \end{pmatrix}$, $P$ representing $f|_V$ and $R$ representing $\tilde{f}$. Let the order $P(f)$ of $f$ be $a$. Then $P^a = I$, $R^a = I$. Let $b$ and $c$ be the orders of $P$ and $R$, respectively. Then $b|a$, $c|a$, so with $d = \text{lcm}(b, c)$ we may put $a = td$. Putting $A^d = \begin{pmatrix} I & S \\ 0 & I \end{pmatrix}$ we have $I = A^{td} = \begin{pmatrix} I & tS \\ 0 & I \end{pmatrix}$, so $S = 0$. This implies $a|d$, so $a = d$. Moreover one may conclude that, if $P$ and $R$ have finite orders $b$ and $c$ respectively, and $d = \text{lcm}(b, c)$, $A$ has finite order if and only if $S = \sum_{i=0}^{d-1} P^i Q R^{d-1-i} = 0$, and

the order of $A$ then is $d$.

Also solved by F. S. Cater, W. Cordwell, S. L. Davis, E. R. Gentile (Argentina), V. D. Mascioni (student, Switzerland), D. M. Wells, P. Y. Wu (Republic of China), and the proposer. Cater generalized the result to the context of a torsion-free Abelian group.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by November 30, 1985. The solver's full post-office address should be on each sheet.*

6498. *Proposed by I. J. Schoenberg, University of Wisconsin, Madison.*

Let $P_0(x) \equiv 1$, $P_1(x)$, $P_2(x), \ldots$ be an Appell sequence of polynomials, i.e.,

$$P_n'(x) = P_{n-1}(x) \qquad (n = 1, 2, \ldots).$$

If $x_{i,n}$ $(i = 1, 2, \ldots, n)$ are the zeros of $P_n(x)$, show that

$$\lim_{n \to \infty} \max_i |x_{i,n}| = \infty,$$

unless $P_n(x) = (x - \alpha)^n / n!$ $(n = 0, 1, \ldots)$ for some fixed $\alpha$.

## SOLUTIONS OF ADVANCED PROBLEMS

### On Equal Integrals

6440 [1983, 569]. *Proposed by M. S. Klamkin, J. McGregor, and A. Meir, University of Alberta.*

Let $F(x)$, $G(x)$ be two functions in $L_1(-\infty, \infty)$ which satisfy

$$\int_{-\infty}^{\infty} F(x)\, dx = \int_{-\infty}^{\infty} G(x)\, dx = 1.$$

Show that for any $\lambda$ in $(0, 1)$ there is a set $E \subseteq (-\infty, \infty)$ such that

$$\int_E F(x)\, dx = \int_E G(x)\, dx = \lambda.$$

*Solution* 1 *by Gerald A. Edgar, The Ohio State University.* By Liapunoff's theorem [see W. Rudin, *Functional Analysis*, second edition, McGraw-Hill, Theorem 5.5], the set $S$ in $\mathbb{R}^2$ defined by

$$S = \left\{ \left( \int_E F(x)\, dx, \int_E G(x)\, dx \right) : E \text{ measurable} \right\}$$

is convex. Since $(0, 0) \in S$ and $(1, 1) \in S$, it follows that $(\lambda, \lambda) \in S$ for any $\lambda \in (0, 1)$.

*Solution* 2 *by the proposers.* (This solution is more elementary than the above.) Let $A = \{x\colon F(x) > G(x)\}$, $B = \mathbb{R} \setminus A$, let $A_t = A \cap (-\infty, t)$, $B_t = B \cap (-\infty, t)$ and let

$$\alpha(t) = \int_{A_t} (F(x) - G(x))\, dx, \qquad \beta(t) = \int_{B_t} (G(x) - F(x))\, dx.$$

Then $\alpha(t)$ and $\beta(t)$ are monotone non-decreasing, continuous functions with $\alpha(-\infty) = \beta(-\infty)$ and $\alpha(\infty) = \beta(\infty)$. Therefore, for every $t$ there exists $\gamma = \gamma(t)$ such that $\alpha(t) = \beta(\gamma)$. Now set $C_t = A_t \cup B_\gamma$. Then clearly $C_t \subset C_s$ when $t \leqslant s$, and $\mu(C_s) \to \mu(C_t)$ when $s \to t+$. Hence the

function

$$\int_{C_t} F(x)\,dx = \int_{C_t} G(x)\,dx =: H(t)$$

is continuous. Since $C_{-\infty} = \varnothing$ and $C_{\infty} = \mathbb{R}$, it follows that $H(t)$ attains all values between 0 and 1. Thus, for $\lambda \in (0,1)$, there is a number $\tau$ such that $H(\tau) = \lambda$ and $E = C_\tau$ has the required property.

Also solved by A. N. Al-Hussaini (Canada), Preben Alsholm & Michael Ratliff, Miroslav D. Ašić (Yugoslavia), Ethan Bolker, Aage Bondesen (Denmark), Victor Hernández (Spain), L. E. Mattics, J. G. Mauldon, Stephen Noltie, M. Pachter (South Africa), Nicholas Passell, Vincent C. Peck, Rae Michael Shortt, and Alfonso Villani (Italy).

### Functionally Countable Spaces

6444 [1983, 648]. *Proposed by Fred Galvin, University of Kansas.*

A topological space $X$ is *functionally countable* if every continuous real-valued function defined on $X$ has a countable range. Prove or disprove: the product of two functionally countable spaces is functionally countable.

*Solution by Salvador Hernández, Facultad de Ciencias Matemáticas, Valencia, Spain.* Let $X = W(\omega_1)$ be the space of all ordinals less than the first uncountable ordinal. Let $Y = D \cup \{p\}$ where $D$ is an uncountable discrete space, $p \notin D$ and, if $p \in V$, then $V$ is a neighborhood of $p$ when the cardinality of $D - V$ is countable. It is clear that $X$ and $Y$ are functionally countable spaces.

Let $\phi$: $X \to D$ be an injective map, and let $H = \{(\alpha, \phi(\alpha))$: $\alpha$ is an isolated point of $X\}$. Then $H$ is an uncountable discrete closed and open subspace of $X \times Y$. Thus $X \times Y$ is not functionally countable.

Also solved by Klaas Pieter Hart (The Netherlands) and the proposer.

### The Generalized Totient Function

6446 [1983, 709–710]. *Proposed by Paul S. Bruckman, Carmichael, CA.*

Define the generalized totient function $\phi_r$ as follows:

$$\phi_r(n) = \sum_{\substack{k=1\\(k,n)=1}}^{n} k^r, \quad r = 0,1,2,\ldots,$$

and its (Dirichlet) generating function

$$f_r(s) = \sum_{n=1}^{\infty} \phi_r(n)/n^s,$$

defined for an appropriate domain of $s$. Show that

$$f_r(s) = \frac{1}{(r+1)\zeta(s-r)} \sum_{k=1}^{r+1} \binom{r+1}{k} |B_{r+1-k}| \zeta(s-k), \quad s > r+2,$$

where the $B_j$'s are Bernoulli numbers and $\zeta$ is the Riemann Zeta function.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* The expression for $f_r(s)$ as stated is incorrect. It should read

$$f_r(s) = 1 + \frac{1}{(r+1)\zeta(s-r)} \sum_{k=1}^{r+1} \binom{r+1}{k} B_{r+1-k} \zeta(s-k).$$

To see this, note that since $\phi_r(n) = O(n^{r+1})$ the series defining $f_r(s)$ is absolutely convergent

when $\operatorname{Re} s > r + 2$ as also is the series defining $\zeta(s - r)$. Hence, for $\operatorname{Re} s > r + 2$,

$$f_r(s)\zeta(s-r) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where

$$a_n = \sum_{d|n} \phi_r(d)\left(\frac{n}{d}\right)^r = n^r \sum_{d|n} \sum_{\substack{k=1\\(k,d)=1}}^{d} \left(\frac{k}{d}\right)^r,$$

$$= n^r \sum_{j=1}^{n} \left(\frac{j}{n}\right)^r = \sum_{j=1}^{n} j^r,$$

and it is known (see, e.g., I. S. Gradshteyn and I. W. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, 1980, 0.121, p. 1) that

$$\sum_{j=1}^{n} j^r = n^r + \frac{1}{r+1} \sum_{k=1}^{r+1} \binom{r+1}{k} B_{r+1-k} n^k.$$

It follows that, for $\operatorname{Re} s > r + 2$,

$$f_r(s)\zeta(s-r) = \zeta(s-r) + \frac{1}{r+1} \sum_{k=1}^{r+1} \binom{r+1}{k} B_{r+1-k} \zeta(s-k).$$

Also solved by C. Georghiou (Greece), Jonathan M. Jacobs, H. Jager (The Netherlands), William Newcomb, and the proposer.

### Factors of $\binom{n}{k}$

6447 [1983, 710]. *Proposed by P. Erdős, Hungarian Academy of Sciences, and J. L. Selfridge, Mathematical Reviews.*

For fixed $k \geqslant 2$, if $n \geqslant 2k$, show that there is at least one $i$, $0 \leqslant i \leqslant k-1$, such that $n - i$ does not divide $\binom{n}{k}$. On the other hand, there is an $n_k \geqslant 2k$ for which $n - i$ divides $\binom{n}{k}$ for all but one $i$ in the range $0 \leqslant i \leqslant k - 1$. Estimate the smallest such $n_k$ as well as you can.

*Partial solution by Jean-Marie Monier, Lyon, France.* Let $k \geqslant 2$, $n \geqslant 2k$. Suppose $n$, $n-1, \ldots, n-k+1$ all divide $\binom{n}{k}$. Since $n(n-1)\cdots(n-k+1) = k!\binom{n}{k}$, we have:

$$\begin{gathered}
(n-1)(n-2)(n-3)\cdots(n-k+1) \in k!\mathbb{N}\\
n(n-2)(n-3)\cdots(n-k+1) \in k!\mathbb{N}\\
\vdots\\
n(n-1)(n-2)\cdots(n-k+2) \in k!\mathbb{N}.
\end{gathered}$$

Subtracting each expression from the one below we get:

$$\begin{gathered}
(n-2)(n-3)(n-4)\cdots(n-k+1) \in k!\mathbb{N}\\
n(n-3)(n-4)\cdots(n-k+1) \in k!\mathbb{N}\\
\vdots\\
n(n-1)(n-2)\cdots(n-k+3) \in k!\mathbb{N}.
\end{gathered}$$

Repeating the process $k-1$ times we arrive at

$$2 \cdot 3 \cdot 4 \cdots (k-1) \in k!\mathbb{N},$$

which is a contradiction since $1/k \notin \mathbb{N}$. Hence there is at least one $i$, $0 \leqslant i \leqslant k-1$, such that $n-i$ does not divide $\binom{n}{k}$.

For the next part of the problem, we observe that for $n_2 = 4$, $n_k = k!$ when $k \geqslant 3$, $\binom{n_k}{k}$ is divisible by $n_k - 1, n_k - 2, \dots, n_k - k + 1$.

Also partially solved by Kee-Wai Lau (Hong Kong) and the proposers. The question of estimating the smallest $n_k \geqslant 2k$ remains open.

**On Clausen's Integral**

6448 [1984, 59]. *Proposed by Henry E. Fettis, Mountain View, CA.*

Show that

$$2\sqrt{3}\,\mathrm{Cl}_2\left(\frac{\pi}{3}\right) = \psi'\left(\frac{1}{3}\right) - \frac{2\pi^2}{3},$$

where $\mathrm{Cl}_2(\theta) = -\int_0^\theta \ln(2\sin\frac{1}{2}t)\,dt$ is Clausen's integral and $\psi(z) = (d/dz)\ln\Gamma(z)$.

*Solution by William A. Newcomb, Lawrence Livermore National Laboratory, Livermore, California.* Let

$$S_1 = \sum_{n=0}^{\infty} \frac{1}{(3n+1)^2}, \qquad S_2 = \sum_{n=0}^{\infty} \frac{1}{(3n+2)^2},$$

$$A_1 = \sum_{n=0}^{\infty} \frac{(-1)^n}{(3n+1)^2}, \qquad A_2 = \sum_{n=0}^{\infty} \frac{(-1)^n}{(3n+2)^2}.$$

Then

$$S_1 + S_2 = \sum_{n=1}^{\infty} \frac{1}{n^2} - \sum_{n=0}^{\infty} \frac{1}{(3n+3)^2} = \frac{\pi^2}{6} - \frac{1}{9}\cdot\frac{\pi^2}{6} = \frac{4\pi^2}{27},$$

$$S_1 - A_1 = 2\sum_{n=0}^{\infty} \frac{1}{(6n+4)^2} = \frac{1}{2}S_2,$$

and

$$S_2 + A_2 = 2\sum_{n=0}^{\infty} \frac{1}{(6n+2)^2} = \frac{1}{2}S_1.$$

Hence

$$A_1 + A_2 = \frac{3}{2}(S_1 - S_2) = 3S_1 - \frac{2\pi^2}{9}.$$

We also have the well-known series representations

$$\mathrm{Cl}_2(\theta) = \sum_{n=1}^{\infty} \frac{\sin n\theta}{n^2}, \qquad 0 \leqslant \theta \leqslant \pi,$$

$$\psi'(t) = \sum_{n=0}^{\infty} \frac{1}{(n+t)^2}, \qquad t \neq 0, -1, -2, \dots .$$

Consequently

$$\psi'(1/3) = 9S_1 \quad \text{and} \quad \mathrm{Cl}_2(\pi/3) = \frac{\sqrt{3}}{2}(A_1 + A_2) = \frac{1}{2\sqrt{3}}\left(\psi'(1/3) - \frac{2\pi^2}{3}\right).$$

Also solved by K. F. Andersen (Canada), Thomas Archibald (Canada), Paul F. Byrd, László Cseh & Imre

Merényi (Romania), C. Georghiou (Greece), M. L. Glasser, O. P. Lossers (The Netherlands), Armel Mercier (Canada), Roger B. Nelsen, Norbert Ortner (Austria), Stan Philipp, Bertram Ross, Lajos Takács, A. Tissier (France), Michael Vowe (Switzerland), and the proposer.

### A Property of Polynomials

6449 [1984, 59]. *Proposed by Kent D. Boklan (student), Massachusetts Institute of Technology.*

Determine the smallest closed subinterval $[a, b]$ of $[0,1]$ such that if $p$ is a polynomial of degree at most 6 that increases on $[a, b]$, then necessarily $p(1) \geqslant p(0)$.

*Solution by Miroslav D. Ašić, University of Belgrade, Yugoslavia.* The smallest such subinterval $[a, b]$ is given by

$$a=\frac{1}{2}-\frac{\sqrt{15}}{10}, \qquad b=\frac{1}{2}+\frac{\sqrt{15}}{10}.$$

Suppose first that $p$ is an arbitrary polynomial of degree at most 6 increasing on $[a, b]$. Then

$$p'(a) \geqslant 0, \qquad p'\left(\frac{1}{2}\right) \geqslant 0, \qquad p'(b) \geqslant 0.$$

Since $\deg p' \leqslant 5$, we may use the corresponding Legendre–Gauss quadrature formula (adapted to the interval $[0,1]$) to obtain

$$p(1)-p(0)=\int_0^1 p'(x)\,dx=\left(5p'(a)+8p'\left(\frac{1}{2}\right)+5p'(b)\right)\Big/18 \geqslant 0.$$

Assume next that there is a shorter subinterval $[c, d] \subseteq [0,1]$ with the required property. Then $a<c$ or $d<b$. Consider the case $a<c$ (the other case can be treated analogously). Let

$$p'(x)=(x-c)\left(x-\frac{1}{2}\right)^2(x-b)^2;$$

obviously $p'(a)<0$ and $p'(x) \geqslant 0$ on $[c, d]$. Using the same quadrature formula again we have

$$p(1)-p(0)=\left(5p'(a)+8p'\left(\frac{1}{2}\right)+5p'(b)\right)\Big/18=\frac{5p'(a)}{18}<0$$

which is a contradiction.

Also solved by Stan Philipp, Doug Tyler & Brad Parsons, and partially solved by the proposer.

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*A Problem Seminar.* By Donald J. Newman. Problem Books in Mathematics, edited by P. R. Halmos. Springer-Verlag, New York, 1982. v + 113 pp.

IVAN NIVEN
*Department of Mathematics, University of Oregon, Eugene, OR 97403*

Here is a challenging book for problem buffs. It is split into three parts, first a list of one hundred and nine problems, some of which have brief introductory statements, then a section giving hints on the author's approaches to problems, and finally the solutions. The problems, at the level of advanced undergraduate work and beginning graduate work, are comparable in difficulty to the problems on the annual Putnam competitions.

What kinds of problems are included in the collection? Some types are identified by the author, as follows: estimation theory, generating functions, expected values in probability, problems on categories in the real variable sense of the word, and convexity. Of the fifty problems of these types, twenty-nine are listed under "estimation theory," by which the author means questions whose answers can be approximated but not expressed in a closed form.

Since the other problems in the book are not classified, we offer here a rough description. There are several problems in each of these areas: asymptotic relations between functions; combinatorial geometry; number theory; real analysis; other probability questions; sequences, especially sequences of integers; partitioning the integers. There are a very few problems in each of these classes: algebraic operations; logic; abstract algebra; maxima and minima; recurrence relations and iterative processes; functional equations; densities of sequences of integers. Many of the problems cross over these lines of classification.

The author's interests are strongly represented, no doubt, in the selection of problems. Thus we find almost no questions in such areas as complex variable theory, geometry apart from combinatorial aspects, graph theory, algebra in general, and numerical analysis. An imbalance of this kind is to be expected in a problem collection by a single author.

How original are the problems? Are they widely known? This reviewer, moderately acquainted with problem collections, already knew eight of the questions. For example, one of the problems was analyzed, with a reference to its source, in an invited lecture at the 1982 summer meetings in Toronto. The only really misplaced problem asks for a proof of one form of Stirling's formula, which is surely a textbook matter. But all in all, the author has assembled a rather novel collection. Moreover, the problems are instructive; many of them illustrate an important point or two. Students can learn a lot of mathematical technique while enjoying the challenges. Furthermore, in many cases a direct attack on a question leads into a morass, so that the book stands as a subtle argument in favor of sophisticated, even abstract, approaches. This is the book's strongest and most attractive feature.

The author is to be commended for doing more in the Solutions section than simply solving the problems. He outlines possible methods of solution, including occasionally approaches that fail. This is an attractive instructional device.

We have three examples to illustrate the author's approaches to problems. Here is problem #18:

$$\text{Evaluate } \sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+\cdots}}}}\,.$$

The author solves this problem by generalizing to the function

$$f(x)=\sqrt{1+x\sqrt{1+(x+1)\sqrt{1+(x+2)\sqrt{1+\cdots}}}}\,,$$

which turns out to be more manageable than the special case with $x=2$.

Next consider problem #76:

$$\text{Prove } n-[n/2]+[n/3]-\cdots\sim n\log 2.$$

Newman points out that removal of the square brackets in $[n/j]$ for $1<j<n$ introduces an error of the same order of magnitude as $n\log 2$, so that this direct approach fails. We must find another approach, perhaps by taking advantage of the alternating signs. (We won't give away the secret here.)

Problem #80 is of a different type:

Given that $f(x)+f'(x)\to 0$ as $x\to\infty$, prove that both $f(x)\to 0$ and $f'(x)\to 0$.

The hint suggests that we should use L'Hôpital's rule on $f(x)e^x/e^x$, since the derivative of $f(x)e^x$ is $\{f(x)+f'(x)\}e^x$.

Turning to the question of who will use the book, the author gives a clear answer when he asks:

"Why have the graduate schools dropped the Problem Seminar?" But it is unlikely that the book will be accorded this role in the already crowded graduate curriculum. Graduate faculty may want their students to solve lots of problems, but not *miscellaneous* problems. They usually want problems directed toward a specific subject, just as R. L. Moore did with his students of topology. Should graduate students spend their time solving such a question as the following, except for spare-time amusement?

> Given any $n$ distinct points in the plane, show that one of the angles determined by them is $\leqslant \pi/n$ (the zero angle counts).

It is often the case that a problem is easy enough for a reader familiar with the topic, but troublesome for others. Consider for example problem #48:

> Call an integer square-full if each of its prime factors occurs to the second power (at least). Prove that there are infinitely many pairs of consecutive square-fulls.

Any student familiar with the elementary theory of the Pell equation knows that for any positive integer $k$ not a perfect square the equation $x^2 - ky^2 = 1$ has infinitely many solutions in integers $x$ and $y$. From $x^2 - 8y^2 = 1$ we get the consecutive square-fulls required. (The author gives a different solution, using only college algebra and a dash of Euler's kind of ingenuity.)

Graduate work aside, the book will find a place in advanced *undergraduate* seminars, such as those preparing students for the Putnam exams. More generally, anyone who likes the challenge of novel problems will enjoy this collection immensely.

---

*Formal Number Theory and Computability*: *A Workbook*. By Alec Fisher. Oxford Logic Series Number 7. Oxford University Press, New York, 1982. xii + 190 pp.

MARTIN D. DAVIS
*Courant Institute, New York University, 251 Mercer Street, New York, NY 10012*

How much should be left to the reader? A tradition which goes back to Euclid calls for a mathematical expositor to include all details. Perhaps the purest example in this tradition in this century is the "Satz-Beweis" style of Edmund Landau. But there are very successful textbooks written with an opposite point of view. The example that comes to mind is Pólya and Szegö's famous two-volume collection of problems in which readers are encouraged to obtain their own proofs, before consulting the carefully arranged outlines of solutions at the back.

The question of which details may be "left to the reader" is particularly troublesome for writers of textbooks in mathematical logic. The methods of mathematical logic require that various processes of ordinary mathematics be "formalized." This means that it needs to be demonstrated that certain proofs and algorithms can be developed within formal systems which, at least superficially, seem severely limited in scope and power. This presents no problem at all for working logicians who have a perfectly clear idea of what can be done using a particular formalism. But it does present a real dilemma for logicians attempting to teach their subject. To follow the Euclid-Landau model and simply present all details is to risk drowning the novice in a sea of tedious details while helping to perpetuate the myth (all too commonly believed by mathematicians a generation ago) that mathematical logic *consists* of nothing but these tedious details. On the other hand, to airily dismiss such details as "obvious" is to risk having the subject lose its cogency, and to leave students with little intuition for what can be formalized in a given system. Alec Fisher's little "workbook" presents an approach which is much closer to Pólya-Szegö than to Euclid-Landau. In only 135 pages of lucid exposition he obtains Gödel's incompleteness theorem (actually the somewhat stronger Gödel-Rosser theorem) for Peano arithmetic. En route

he obtains Presburger's decision procedure for arithmetic without multiplication and, in addition, he shows that the problem of deciding whether a given computational procedure will ever halt is algorithmically undecidable. Many details are left as exercises, and fairly detailed outline solutions are presented in an additional 50 pages. The book is intended for quite unsophisticated readers, is very clearly written, and should provide a very effective teaching tool.

In order to accomplish his task, Fisher required not only a logical formalism for elementary number theory, but also a model of computation for giving a precise definition of *computable* function. From the many mathematically equivalent formalisms available for the purpose, Fisher chose "register" machines. This simple and elegant formalism is highly accessible, especially to readers with a modicum of programming experience. However, because Fisher is working with a formalization of number theory in which $+$ and $\cdot$ are the only functions explicitly available, the task of proving that all computable functions are representable (in a suitable sense) in his formalization is rather formidable. Fisher uses the now classical device, Gödel's $\beta$ function:

$$\beta(c,d,i) = \mathrm{rm}(c, 1 + (i+1)d),$$

where $\mathrm{rm}(x, y)$ is the least nonnegative remainder when $x$ is divided by $y$. Using the Chinese remainder theorem, it is not too difficult to show that for every finite sequence $a_0, a_1, \dots, a_n$, there are integers $c, d$ such that $\beta(c,d,i) = a_i$ for $i = 0, 1, \dots, n$. The successive steps of a computation by a register machine program can then be coded using this $\beta$ function. The catch is that it is not much fun to formalize this in Peano arithmetic.

An alternative path, which avoids much of the tedious complications, is first to prove as a theorem of ordinary mathematics that computability by register machines is equivalent to a more tractable notion. For example, by a theorem of Kleene (later improved by Julia Robinson), easily proved by an *informal* use of Gödel's $\beta$ function, the computable functions are just those obtainable from a few initial functions using the two operations of function composition, and taking the least zero of a function. Another attractive possibility only recently available is to give the new Jones-Matijasevič proof of the Davis-Putnam-Robinson theorem that the graph of a computable function consists of the tuples which satisfy an exponential Diophantine equation. This new proof is very simple and is carried out using register machines. If this path is taken, the exponential function should probably be added to $+$ and $\cdot$ as primitive functions for which the recursion equations are taken as axioms. The slight weakening of the Gödel-Rosser theorem that would result hardly matters at this level, and as a dividend one would obtain the unsolvability of the decision problem for exponential Diophantine equations. (Alternatively, one could digress to obtain Matijasevič's theorem that the exponential function is Diophantine, thereby not only avoiding the weakening of the Gödel-Rosser theorem, but also obtaining the unsolvability of Hilbert's tenth problem.)

Two minor errata: On p. 34 it is stated that Russell and Whitehead's *Principia Mathematica* was written independently of Frege's work, which is certainly not the case. On p. 37 it is stated that there is no mechanical procedure for "deciding whether, for arbitrary $n$, a sequence of $n$ successive 7s occurs in the decimal expansion of $\pi$, ... ." What is presumably meant is that no such procedure is *known*.

---

**ANSWERS TO REBUSES ON PAGES 426–427**

1. Triangle inequality.
2. The inverse of a matrix.
3. Iceberg (= ice cube + sea = (ice)$^3$ + C).

4a. Power series.

4b. Formal power series.

5. Telescoping series.
6. Subset.

*Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems*. By Richard Haberman. Prentice-Hall, Englewood Cliffs, NJ, 1983. x + 533 pp. $34.95.

Gregory B. Passty
*Department of Mathematics and Computer Science, Southwest Texas State University, San Marcos, TX 78666*

Whenever I have taught Fourier series and their applications, I have found two topics that merit special care—the modeling process and the validity of series solutions. While inclusion of both topics is necessary for a complete course in applications of Fourier series with mathematical rigor, either could be considerably deemphasized without seriously affecting the rest of the material. Testimony to this is provided by the varying treatments given in the different textbooks used to teach the subject today.

Two basic types of students take such a course. One group is made up of mathematics majors looking for an introduction to partial differential equations. The other group comprises science majors, most notably from physics and engineering, who are looking for mathematical understanding of problems and problem-solving techniques that they may have already encountered in their own majors.

The mathematics majors should have more of an interest in the difference between formal and actual solutions and associated ideas—convergence, term-by-term differentiation, etc. They seem to have less interest in the modeling process because they haven't really practiced it very much, and also because, with several recently completed abstract courses behind them, they have yet to perceive how thoroughly involved with modeling some mathematical careers can be. For the science majors, the steps and decisions leading to partial differential equations that describe natural phenomena and are, at the same time, amenable to analysis (if not complete solution) seem far more important than a discussion of validity. They know that the methods work in situations of interest to them, anyway.

With *Elementary Applied Partial Differential Equations*, Richard Haberman presents a very thorough text for a first course in partial differential equations while finding excellent approaches to the two topics I have mentioned. The first half of the book deals with the method of separation of variables and all that entails. Following a chapter on nonhomogeneous problems, the second half comprises several other topics: Green's functions, the method of characteristics, Fourier and Laplace transforms, and finite difference numerical methods.

This book will appeal to both of the audiences mentioned above. Haberman begins with a chapter-long derivation of the heat equation and its various boundary conditions in one and then two or three dimensions. All physical ideas are very patiently and clearly explained. Specific heat, for example, is introduced after we are brought to the realization that we may need to consider some distinction between thermal energy and temperature. The development leads first to an integral conservation law and then to the heat equation as we generally think of it. Each step of the process is made clear, and many variations are either explained or suggested. Note that the derivation (via the divergence theorem) of the heat equation in several dimensions is given here in Chapter 1 rather than later, as in most other treatments of the subject I have read.

An important feature of this first chapter which persists throughout the book is the repeated appeal to physical reasoning to suggest or interpret the mathematics. In the exposition and also in the exercises, the student is often asked to explain physically the meaning of a certain mathematical result. Haberman does not use physical reasoning in place of proofs, however. He communicates to the student that mathematics and physics can work closely together, each illuminating the other. In Chapter 1, the modeling concepts are stressed as much as the mathematics, and ideas developed here are repeatedly referred to subsequently, for example to explain in advance why the separation constant is most naturally written with a minus sign.

Chapters 2, 3, and 4 consist of series solutions of the heat equation, wave equation, and potential equation (with various boundary conditions), as well as the required discussion of Fourier series, half-range expansions, and so on. In spite of this array of different problems, there

has appeared through the first four chapters only one differential equation for the spatial eigenfunctions: $\phi'' = -\lambda\phi$, with various boundary conditions. This displays an admirable pedagogical quality: resistance to the temptation to throw in too many new things all at once, especially when they're not needed for the discussion at hand. (Other differential equations do appear in the exercises.) The same quality is manifested in Chapter 7, in a section on forced vibrations of a membrane *of arbitrary shape*. The shape is unspecified, so the corresponding eigenfunctions must go uncomputed. While the problem thus remains fairly general, Haberman is able to carry on a very informative discussion of forcing, periodic forcing, and resonance without cluttering the landscape with a pedagogically unnecessary digression.

The author presents quite a few proofs throughout the book. Most of the time, they are included as needed in the exposition. In some cases (e.g., the Riemann-Lebesgue lemmas), they are left for optional sections or appendices.

We now summarize briefly Haberman's discussion of the validity of the series solution to the heat equation. Let $f(x)$ be a periodic function. The keys are the following two theorems:

(1) [p. 99]. *If $f'(x)$ is piecewise smooth, then a continuous Fourier series corresponding to $f(x)$ can be differentiated term by term.*

(2) [p. 93]. *For a piecewise smooth $f(x)$, the Fourier series corresponding to $f(x)$ is continuous if and only if $f(x)$ is continuous.*

Now consider the validity of

$$u(x,t) = \sum_{n=1}^{\infty} B_n \sin\frac{n\pi x}{L} \exp\left(-\frac{n^2\pi^2}{L^2}kt\right)$$

as a solution to the heat equation $u_t = ku_{xx}$ with boundary conditions $u(0,t) = u(L,t) = 0$, and initial condition $u(x,0) = f(x)$. The $B_n$'s are of course the Fourier sine series coefficients for $f(x)$. Under the assumption that $u(x,t)$ and $\partial u/\partial x(x,t)$ are continuous functions of $x$ (required by the physical formulation) and that $\partial^2 u/\partial x^2$ is at least piecewise smooth, we are justified in computing the second derivative $\partial^2 u/\partial x^2$ term by term. On the other hand, term-by-term differentiation with respect to $t$ is justified if $u(x,t)$ is continuous and $\partial u/\partial t$ is piecewise smooth. (The proof, depending on Leibniz's rule, is left as an exercise.) Comparison of the term-by-term derivatives shows the heat equation is satisfied by $u(x,t)$.

Relying upon the continuity assumption, Haberman never mentions the notion of uniform convergence, which can be used to derive the regularity properties hypothesized for $u$ and its derivatives from regularity properties of $f(x)$. But this omission is certainly acceptable in a course where most of the students have not encountered uniform convergence in their previous studies.

Looking at other texts, I find that most do not discuss the validity question at all, but do say that the solution obtained is a *formal* solution. Kreyszig [**3**, p. 423ff.] demonstrates a validity argument for the solution of the wave equation (with zero initial velocity for simplicity), and notes that validity ultimately follows from continuity of the initial displacement and piecewise continuity of its derivatives. There is again no mention of uniform convergence in this context.

The first book from which I learned differential equations, by Kreider *et al.* [**2**, p. 380ff., p. 522ff.], does discuss uniform convergence and its consequences for Fourier series. It is pointed out that the initial condition really says $\lim_{t\to 0^+} u(x,t) = f(x)$ for all $x$ in $(0,L)$, and that uniform convergence supplies the continuity necessary to verify that this really happens. Churchill [**1**] also shows how the necessary continuity properties follow from regularity of the initial and/or boundary conditions by way of Abel's Test for Uniform Convergence. I feel that both of these very good treatments would be difficult to follow for a class without earlier experience with uniform convergence. Haberman's compromise is a good one.

Problems treated in the text are almost all variants of the heat, wave, or potential equation, with various nonhomogeneities. Quite a number of other variations appear in the exercises. There is some discussion of non-physical problems as well. These are used to shed additional light on Sturm-Liouville eigenvalue problems, the subject of Chapter 5.

A sufficient prerequisite for the student is ordinary differential equations, although some later sections suggest that prior brief exposures to complex variables and to Laplace transforms would be helpful.

Haberman's book is a fine treatment of elementary partial differential equations. I recommend it for anyone who wants to illustrate to math and science students the ways in which their disciplines interact and the insights each subject contributes in elucidating the other.

**References**

1. Ruel V. Churchill, Fourier Series and Boundary Value Problems, 2nd ed., McGraw-Hill, New York, 1969.
2. Donald L. Kreider, Robert G. Kuller, Donald R. Ostberg, and Fred W. Perkins, An Introduction to Linear Analysis, Addison-Wesley, Reading, MA, 1966.
3. Erwin Kreyszig, Advanced Engineering Mathematics, 3rd ed., Wiley, New York, 1972.

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

Miscellaneum 129 ("Triangles are square," June-July 1984 MONTHLY) may have misled many readers. Here is some background on the item.

That $n^2$ points fall naturally into a triangular array is a not-quite-obvious fact which may have applications (e.g., to symmetries of Latin-square "$k$-nets") and seems worth stating more formally. To this end, call a convex polytope $P$ an *$n$-replica* if $P$ consists of $n$ mutually congruent polytopes similar to $P$ packed together. Thus, for $n \in \mathbb{N}$,

(A) An equilateral triangle is an $n$-replica if and only if $n$ is a square.

Does this generalize to tetrahedra, or to other triangles? A regular tetrahedron is not a $(2^3)$-replica, but a tetrahedron $ABCD$ with edges $AB$, $BC$, and $CD$ equal and mutually orthogonal is an $n$-replica if and only if $n$ is a cube. Every triangle satisfies the "if" in (A), so, letting $T$ be the set of triangles, one might surmise that

(B) $\forall t \in T$ ($t$ is an $n$-replica if and only if $n$ is a square).

This, however, is false. A. J. Schwenk has pointed out that for any $m \in \mathbb{N}$, the 30°-60°-90° triangle is a $(3m^2)$-replica, and that a right triangle with legs of integer lengths $a$ and $b$ is an $((a^2 + b^2)m^2)$-replica. As Schwenk notes, it does not seem obvious which other values of $n$ can occur in counterexamples to (B). Shifting parentheses to fix (B), we get a "square-triangle" lemma:

(C) ($\forall t \in T$, $t$ is an $n$-replica) if and only if $n$ is a square.

Miscellaneum 129 was a less formal statement of (C), with quotation marks instead of parentheses; this may have led many readers to think (B) was intended. To these readers, my apology.

Steven H. Cullinane
501 Follett Run Road
Warren, PA 16365

Editor:

Professor Neal Koblitz has pointed out to me that there is an error in my paper "Factorization and primality tests" (this MONTHLY, 91 (1984) 333–352) in the criterion for irreducibility stated in Exercise 9. The criterion should read:

A monic polynomial $f(X)$ of degree $d$ over $\mathbf{Z}_p$ is irreducible if and only if
(i) $X^{p^d} \equiv X(\operatorname{mod} f(X))$; but
(ii) $\operatorname{GCD}(X^{p^{d/q}} - X, f(X)) = 1$ for each prime $q|d$.

Dr. Gustavus J. Simmons has sent me a report ("Status Report on Factoring at the Sandia National Laboratory") about the work done on factorization by himself and colleagues J. A. Davis and Diane B. Holdridge at Sandia National Laboratories, Albuquerque, New Mexico, during the past few years. The report describes a method used for fairly routine factorization of numbers up to about 70 digits (as compared to the figure of 50 digits given in my paper.) This recent advance is credited to three things: use of Carl Pomerance's quadratic sieve (an improvement on the continued fraction method) with a technical modification by J. A. Davis which makes it run particularly fast, the use of a very fast Cray computer, and programming by Tony Warnock and Diane Holdridge which takes full advantage of the parallel architecture of the Cray. The Cray computers form one of several families of machines which are capable of doing thousands of computations simultaneously (in "parallel"), but special methods of programming are required in order to take full advantage of this property. Undoubtedly further advances in hardware will allow faster factoring programs even if no better algorithms are used.

John D. Dixon
Department of Mathematics
Carleton University
Ottawa, Ontario K1S 5B6
Canada

---

## 150. MISCELLANEA

### Is Mathematics a Mystery?

As evidence goes, I favor what is usually called "circumstantial evidence" as against the testimony of witnesses. . . . Neither have I much faith in experts who claim infallibility in any field except, perhaps, abstract mathematics.

—*Dashiell Hammett*, by
Diane Johnson, Random House,
New York, 1983, p. 18.

It would be impossible now for me to remember all that he wanted to learn, but I remember a long year of study on the retina of the eyes; how to play chess in your head; the Icelandic sagas; the history of the snapping turtle; Hegel; would a hearing aid—he bought a very good one—help in detecting bird sounds; then from Hegel, of course, to Marx and Engels straight through; the shore life of the Atlantic; and finally, and for the rest of his life, mathematics. He was more interested in mathematics than in any other subject except baseball.

—Lillian Hellman, from *Dashiell Hammett*, by Diane Johnson,
Random House, New York, 1983, p. 169.



THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 7 August-September 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.




The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# INTERESTING SERIES INVOLVING THE CENTRAL BINOMIAL COEFFICIENT

D. H. LEHMER
*Department of Mathematics, University of California, Berkeley, CA 94720*

The adjective "interesting" is used here in a technical sense explained as follows. A series is called interesting in case there is a simple explicit formula for its $n$th term and at the same time its sum can be expressed in terms of known constants. Thus

$$1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n}+\cdots=2,$$

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots+\frac{(-1)^{n+1}}{n}+\cdots=\log 2,$$

$$1+\frac{1}{16}+\frac{1}{81}+\frac{1}{256}+\cdots+\frac{1}{n^4}+\cdots=\frac{\pi^4}{90},$$

are familiar examples of interesting series.

The series we plan to discuss are of two types:

$$\text{I.}\quad \sum_{n=0}^{\infty} a_n\binom{2n}{n} \qquad \text{and} \qquad \text{II.}\quad \sum_{n=0}^{\infty}\frac{a_n}{\binom{2n}{n}},$$

where the $a_n$ are very simple functions of $n$. We begin with series of Type I.

By the binomial theorem we have

$$\sum_{n=0}^{\infty}\binom{2n}{n}x^n=\frac{1}{\sqrt{1-4x}}. \tag{1}$$

This converges if $|x|<1/4$. If we put $x=1/8$, for example, we get the interesting series

$$1+\frac{1}{4}+\frac{3}{32}+\frac{5}{128}+\cdots+\frac{\binom{2n}{n}}{8^n}+\cdots=\sqrt{2}. \tag{2}$$

If $x=1/10$, we get

$$1+.2+.06+.02+.007+\cdots+\frac{\binom{2n}{n}}{10^n}+\cdots=\sqrt{\frac{5}{3}}. \tag{3}$$

For $x=-1/8$

$$1-\frac{1}{4}+\frac{3}{32}-\frac{5}{128}+\cdots+\frac{(-1)^n\binom{2n}{n}}{8^n}+\cdots=\sqrt{\frac{2}{3}}. \tag{4}$$

Averaging (2) and (4) gives us

$$\sum_{n=0}^{\infty}\frac{\binom{4n}{2n}}{64^n}=\frac{3\sqrt{2}+\sqrt{6}}{6}.$$

Another step along this path is to bring in complex variables. Thus using $ix$ instead of $x$ leads to

*D. H. Lehmer*: This paper marks the sixtieth anniversary of the publication of the author's first paper in this MONTHLY. The author hopes to submit manuscripts of other papers from time to time as occasion arises.

$$4\sum_{n=0}^{\infty}\frac{\binom{8n}{4n}}{8^{4n}}=\frac{3\sqrt{2}+\sqrt{6}}{3}+\frac{2\sqrt{2+\sqrt{5}}}{\sqrt{5}},$$

or

$$\sum_{n=0}^{\infty}\frac{\binom{8n}{4n}}{8^{4n}}=\frac{\left(15\sqrt{2}+5\sqrt{6}+6\sqrt{5}\sqrt{2+\sqrt{5}}\right)}{60}.$$

In general one can obtain a value for the sum

$$\sum_{n=0}^{\infty}\frac{\binom{2(an+b)}{an+b}}{c^{an+b}}$$

by setting $x=\varepsilon^{\nu}/c$ $(\nu=0,1,2,\ldots,a-1)$, where $\varepsilon=\exp(2\pi i/a)$, in (1) and forming the appropriate linear combination of these results, a process known as multisection (see Riordan [**4**]).

Another path out of (1) is to apply operators. If we integrate (1) from 0 to $x$ we obtain

$$\int_0^x\sum_{n=0}^{\infty}\binom{2n}{n}t^n\,dt=\int_0^x\frac{dt}{(1-4t)^{1/2}}.$$

Dividing both sides by $x$ we get

$$\sum_{n=0}^{\infty}\binom{2n}{n}\frac{x^n}{n+1}=\frac{1}{2x}\left(1-\sqrt{1-4x}\right). \tag{5}$$

The coefficients $C_n$ of the series are integers known as Catalan numbers. Their first ten values are

$$1,1,2,5,14,42,132,429,1430,4862,\ldots$$

and they occur frequently in combinatorial analysis (see Gould [**2**]). By using Stirling's formula for $n!$ we see that

$$C_n\sim\frac{4^n}{\sqrt{\pi n}\,(n+1)}\left(1-\frac{1}{8n}\right).$$

From this we see that for $x=1/4$ the series in (5) converges slowly like $\Sigma(1/n^{3/2})$. Thus (5) gives us the interesting series

$$1+\frac{1}{4}+\frac{1}{8}+\frac{5}{64}+\cdots+\frac{\binom{2n}{n}}{4^n(n+1)}+\cdots=2.$$

We can treat (1) a little differently by transposing its first term to the right side, dividing both sides by $x$ and then integrating. This gives us the improper integral

$$\int_0^x\sum_{n=1}^{\infty}\binom{2n}{n}t^{n-1}\,dt=\int_0^x\left(\frac{1}{t(1-4t)^{1/2}}-\frac{1}{t}\right)dt.$$

That is,

$$\sum_{n=1}^{\infty}\frac{1}{n}\binom{2n}{n}x^n=2\log\left(\frac{1-\sqrt{1-4x}}{2x}\right). \tag{6}$$

Putting $x=\frac{1}{4}$ we get the series

$$\frac{1}{2}+\frac{3}{16}+\frac{5}{48}+\frac{35}{512}+\cdots+\frac{\binom{2n}{n}}{n4^n}+\cdots=\log 4.$$

If we alternate the signs of this series, we get

$$\frac{1}{2}-\frac{3}{16}+\frac{5}{48}-\frac{35}{512}+\cdots+\frac{(-1)^{n+1}\binom{2n}{n}}{n4^n}+\cdots=\log\left[\frac{\sqrt{2}+1}{2}\right].$$

Adding or subtracting these two series we can derive interesting series for the two sums

$$\sum_{n\text{ odd}}\binom{2n}{n}\frac{1}{n4^n}\qquad\text{or}\qquad\sum_{\substack{n\text{ even}\\ n>0}}\binom{2n}{n}\frac{1}{n4^n}.$$

We can integrate (6) and obtain

$$x\sum_{n=1}^{\infty}\frac{\binom{2n}{n}}{n(n+1)}x^n=2x\log\frac{1-\sqrt{1-4x}}{x}+\frac{\sqrt{1-4x}}{2}+x(\log 4-1)-\frac{1}{2}. \tag{7}$$

If we substitute $x=1/4$, we obtain the interesting series

$$\frac{1}{4}+\frac{1}{16}+\frac{5}{192}+\frac{7}{512}+\cdots+\frac{\binom{2n}{n}}{n(n+1)4^n}+\cdots=\log 4-1.$$

Any number of examples of this sort can be obtained by repeated integrations of (1).

Another operator of this kind is differentiation. Let $\vartheta$ be the operator

$$\vartheta=\frac{xd}{dx}.$$

If we apply $\vartheta$ to (1), we get

$$\sum_{n=1}^{\infty}n\binom{2n}{n}x^n=\vartheta\left(\frac{1}{\sqrt{1-4x}}\right)=2x(1-4x)^{-3/2}. \tag{8}$$

If we set $x=\frac{1}{8}$, we get

$$\frac{1}{4}+\frac{3}{16}+\frac{15}{125}+\frac{35}{512}+\cdots+\frac{n\binom{2n}{n}}{8^n}+\cdots=\frac{\sqrt{2}}{2}.$$

Operating again by $\vartheta$ we obtain

$$\sum_{n=1}^{\infty}n^2\binom{2n}{n}x^n=\vartheta^2(1-4x)^{-1/2}=\frac{2x(2x+1)}{(1-4x)^{5/2}}.$$

Setting $x=\frac{1}{8}$, we find

$$\frac{1}{4}+\frac{3}{8}+\frac{45}{128}+\frac{35}{128}+\frac{1575}{8192}+\cdots+\frac{n^2\binom{2n}{n}}{8^n}+\cdots=\frac{5\sqrt{2}}{4}.$$

If in (1) we replace $x$ by $x^2$ and then integrate both sides, we get

$$\sum_{n=0}^{\infty}\frac{\binom{2n}{n}}{2n+1}x^{2n}=\frac{1}{2x}(\arcsin 2x)$$

and for $x=\frac{1}{4}$

$$\sum_{n=0}^{\infty}\frac{\binom{2n}{n}}{(2n+1)16^n}=\frac{\pi}{3}.$$

We turn now to series of type II which involve the central binomial coefficient in the denominator. These series are more mysterious and less well understood. In an informal

discussion of Apery's proof of the irrationality of $\zeta(3)$, A. J. van der Poorten [5] refers to four series of type II. The following theorem was quoted and its proof by Z. A. Melzak [3] was described as "not quite appropriate". We give an entirely differently proof.

THEOREM. *If* $|x| < 1$,

$$\frac{2x\arcsin x}{\sqrt{1-x^2}} = \sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m\binom{2m}{m}}. \tag{9}$$

*Proof*. We use the familiar Gregory series

$$t\arctan t = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}t^{2m}}{2m-1} \tag{10}$$

and set $t = x/\sqrt{1-x^2}$, so that $\arctan t = \arcsin x$. Then (10) becomes

$$\begin{aligned}\frac{x}{\sqrt{1-x^2}}\arcsin x &= \sum_{m=1}^{\infty} \frac{(-1)^{m-1}x^{2m}}{(2m-1)(1-x^2)^m}\\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{2m-1}\sum_{j=0}^{\infty}(-1)^j x^{2(j+m)}\binom{-m}{j}\\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{2m-1}\sum_{j=0}^{\infty}\binom{m+j-1}{j}x^{2(j+m)}\\ &= \sum_{r=1}^{\infty} x^{2r}\sum_{m=1}^{r}\frac{(-1)^{m-1}(r-1)!}{(m-1)!(r-m)!(2m-1)}.\end{aligned}$$

It suffices to show that the coefficient of $x^{2r}$ is half of that on the right side of (9). That is,

$$r\binom{2r}{r}\sum_{\nu=0}^{r-1}\frac{(-1)^{\nu}(r-1)!}{\nu!(r-\nu-1)!(2\nu+1)} = 2^{2r-1}. \tag{11}$$

To prove this we use Wallis' integral

$$\int_0^{\pi/2}(\sin\theta)^{2r-1}\,d\theta = \frac{2\cdot 4\cdot 6\cdots(2r-2)}{1\cdot 3\cdot 5\cdots(2r-1)}$$

as follows. The left-hand side of (11) can be written

$$\begin{aligned}r\binom{2r}{r}\sum_{\nu=0}^{r-1}(-1)^{\nu}\binom{r-1}{\nu}\frac{1}{2\nu+1} &= r\binom{2r}{r}\int_0^1\sum_{\nu=0}^{r-1}(-1)^{\nu}\binom{r-1}{\nu}y^{2\nu}\,dy\\ &= r\binom{2r}{r}\int_0^1(1-y^2)^{r-1}\,dy\\ &= r\binom{2r}{r}\int_0^{\pi/2}(\sin\theta)^{2r-1}\,d\theta\end{aligned}$$

by substituting $y = \cos\theta$. Using Wallis' integral we find

$$r\binom{2r}{r}\frac{2^{2r-2}(r-1)!(r-1)!}{(2r-1)!} = 2^{2r-1}.$$

This proves (11) and hence the theorem is established.

If we substitute $x = \frac{1}{2}$, we get the result (van der Poorten [5], p. 202)

$$\sum_{m=1}^{\infty} \frac{1}{m\binom{2m}{m}} = \frac{\pi\sqrt{3}}{9} \tag{12}$$

as our first interesting series of Type II. If we wish to alternate the signs in this series, we merely put $x = i/2$ in (9). This gives us

$$\sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m\binom{2m}{m}} = \log\left(\frac{1+\sqrt{5}}{2}\right)/\sqrt{5}.$$

If we divide both members of (9) by $2x$ and then integrate, we obtain

$$2(\arcsin x)^2 = \sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m^2\binom{2m}{m}} \tag{13}$$

from which we get, for $x = \frac{1}{2}$,

$$\sum_{m=1}^{\infty} \frac{1}{m^2\binom{2m}{m}} = \frac{\pi^2}{18} = \frac{1}{3}\zeta(2), \tag{14}$$

where $\zeta(s)$ is Riemann's zeta function. Also,

$$\sum_{m=1}^{\infty} \frac{(-1)^m}{m^2\binom{2m}{m}} = 2\left\{\log\left(\frac{\sqrt{5}+1}{2}\right)\right\}^2$$

If we divide both sides of (13) by $2x$ and integrate, we obtain

$$\sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m^3\binom{2m}{m}} = 4\int_0^x \frac{(\arcsin y)^2}{y}\,dy.$$

This integral is a "higher transcendent". It is closely connected with Spence's transcendent, Clausen's integral, and the trigamma function. For $x = \frac{1}{2}$ we obtain

$$\begin{aligned}\sum_{m=1}^{\infty} \frac{1}{m^3\binom{2m}{m}} &= 4\int_0^{1/2}(\arcsin y)^2\frac{dy}{y}\\ &= -2\int_0^{\pi/3} x\log\left(2\sin\frac{x}{2}\right)dx\\ &= -\frac{\zeta(3)}{3} - \frac{\pi\sqrt{3}}{72}\left\{\psi\left(\frac{1}{3}\right)-\psi\left(\frac{2}{3}\right)\right\},\end{aligned}$$

where $\psi(x)$ is the trigamma function. Van der Poorten [6] rejects this evaluation as being non-instructive. However, he does give the interesting series

$$\frac{1}{2} - \frac{1}{48} + \frac{1}{540} - \frac{1}{4480} + \cdots + \frac{(-1)^{m-1}}{m^3\binom{2m}{m}} + \cdots = \frac{2}{5}\zeta(3)$$

as well as Comtet's [1] remarkable

$$\sum_{m=1}^{\infty} \frac{1}{m^4\binom{2m}{m}} = \frac{17\pi^4}{3240}.$$

There are no known interesting series of the form

$$\sum_{m=1}^{\infty} \frac{1}{m^k \binom{2m}{m}},$$

for $k > 4$.

If we apply the operator $\vartheta^k$ to both sides of (9), we get an almost unlimited number of interesting series.

To begin with $\vartheta$ we obtain

$$\sum_{m=1}^{\infty} \frac{(2x)^{2m}}{\binom{2m}{m}} = \frac{x^2}{1-x^2} + \frac{x \arcsin x}{(1-x^2)^{3/2}}. \tag{15}$$

From this we obtain, for example,

$$\frac{1}{2} + \frac{1}{6} + \frac{1}{20} + \frac{1}{70} + \cdots + \frac{1}{\binom{2n}{n}} + \cdots = \frac{9 + 2\pi\sqrt{3}}{27}.$$

If we replace $x$ by $ix$ in (15) and set $x = a/b$, we get another form of (15)

$$\sum_{m=1}^{\infty} \frac{(-1)^{m-1}(2a/b)^{2m}}{\binom{2m}{m}} = \frac{ab^2}{h^3}\left[\log\left(\frac{a+h}{b}\right) + \frac{ah}{b^2}\right],$$

where $h = \sqrt{a^2 + b^2}$. For example, if $a = 1$, $b = 2$, $h = \sqrt{5}$, we get

$$\frac{1}{2} - \frac{1}{6} + \frac{1}{20} - \frac{1}{70} + \cdots + \frac{(-1)^{n-1}}{\binom{2n}{n}} + \cdots = \frac{1}{5} + \frac{4\sqrt{5}}{25}\log\frac{\sqrt{5}+1}{2}.$$

For $a = 23660$, $b = 23661$, we have $h = 33461$ and we obtain

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{\left(\frac{23660}{23661}\right)^n 4^n}{\binom{2n}{n}} = \frac{23660}{(33461)^2}\left[23660 + \frac{(23661)^2}{33461}\log\frac{57121}{23661}\right]$$

$$= .811587506\ldots.$$

This series is of course convergent. If one examines the ratio of consecutive terms one finds that the terms increase in absolute value until $n = 11830$ when the terms are as large as 117. After that they decrease to zero. It is well that we have our formula for the sum.

If we operate on (9) by higher powers of $\vartheta$, the series we obtain are of the form

$$\sum_{m=1}^{\infty} \frac{m^{k-2}(2x)^{2m}}{\binom{2m}{m}} \qquad (k \geqslant 0) \tag{16}$$

and the value of the sum depends, as a function of $k$, on two sequences of polynomials $V_k(t)$ and $W_k(t)$ defined recursively as follows:

$$V_1(t) = 1, \qquad W_1(t) = 0,$$

$$V_{k+1}(t) = \{(2k-2)t + 1\}V_k(t) + 2(1-t)\vartheta V_k(t), \tag{17}$$

$$W_{k+1}(t) = \{(2k-4)t + 2\}W_K(t) + 2(1-t)\vartheta W_k(t) + V_k(t). \tag{18}$$

The first few polynomials are

$$V_2 = 1, \qquad\qquad W_1 = 1,$$

$$\begin{aligned} V_3 &= 1 + 2t, & W_3 &= 3, \\ V_4 &= 1 + 10t + 4t^2, & W_4 &= 7 + 8t, \\ V_5 &= 1 + 36t + 60t^2 + 8t^3, & W_5 &= 15 + 70t + 20t^2. \end{aligned}$$

The value of the sum (16) can be given explicitly in terms of $V_k$ and $W_k$ by

$$\frac{x}{2^{k-2}(1-x^2)^{k-1/2}}\left[\arcsin x V_k(x^2) + x\sqrt{1-x^2}\, W_k(x^2)\right].$$

If we operate on this by $\vartheta$ and collect the coefficients of arcsin $x$ and $x\sqrt{1-x^2}$, we obtain the formulas (17) and (18). A trigonometric form of this result is, with $x = \sin\theta$,

$$\sum_{m=1}^{\infty} \frac{m^{k-2}4^m(\sin\theta)^{2m}}{\binom{2m}{m}} = \frac{\sin 2\theta}{(2\cos^2\theta)^k}\left[2\theta V_k(\sin^2\theta) + \sin 2\theta W_k(\sin^2\theta)\right].$$

If we replace $x$ by $ix$ we get for the alternating series

$$\sum_{m=1}^{\infty} \frac{(-1)^{m-1}m^{k-2}4^m(\sinh z)^{2m}}{\binom{2m}{m}} = \frac{\sinh 2z}{(2\cosh^2 z)^k}\left[2\log\{\sinh z + \cosh z\}\right.$$

$$\left.\times V_k(-\sinh^2 z) + \sinh 2z W_k(-\sinh^2 z)\right].$$

These two formulas with their two parameters $\theta$ and $k$ yield a wide variety of interesting series as examples. We list only the following. The sums extend from 1 to $\infty$.

$$\sum \frac{m.}{\binom{2m}{m}} = \frac{2}{27}(\pi\sqrt{3} + 9)$$

$$\sum \frac{m^2}{\binom{2m}{m}} = \frac{2}{81}\{5\pi\sqrt{3} + 54\}$$

$$\sum \frac{m^3}{\binom{2m}{m}} = \frac{2}{243}\{37\pi\sqrt{3} + 405\}$$

$$\sum \frac{(-1)^{m-1}m}{\binom{2m}{m}} = \frac{2}{125}[2\sigma + 15] \qquad \left(\sigma = \sqrt{5}\log\frac{1+\sqrt{5}}{2} = 1.076022352\right)$$

$$\sum \frac{(-1)^{m-1}m^2}{\binom{2m}{m}} = \frac{4}{125}[5 - \sigma]$$

$$\sum \frac{(-1)^{m-1}m^3}{\binom{2m}{m}} = \frac{2}{625}[28\sigma + 5]$$

$$\sum \frac{2^m}{m^2\binom{2m}{m}} = \frac{\pi^2}{8}$$

$$\sum \frac{2^m}{m\binom{2m}{m}} = \frac{\pi}{2}$$

$$\sum \frac{2^m}{\binom{2m}{m}} = \frac{\pi}{2} + 1$$

$$\sum \frac{m2^m}{\binom{2m}{m}} = \pi + 3$$

$$\sum \frac{m^2 2^m}{\binom{2m}{m}} = \frac{1}{5} \sum \frac{m^3 2^m}{\binom{2m}{m}} = \frac{7\pi}{2} + 11$$

$$\sum \frac{m^4 2^m}{\binom{2m}{m}} = 113\pi + 355$$

$$\sum \frac{m^{10} 2^m}{\binom{2m}{m}} = 229093376\pi + 719718067.$$

In general,

$$\sum_{m=1}^{\infty} \frac{m^k 2^m}{\binom{2m}{m}} = a\pi + b,$$

where $b/a$ is a close approximation to $\pi$.

$$\sum \frac{3^m}{m^2\binom{2m}{m}} = \frac{2\pi^2}{9}$$

$$\sum \frac{3^m}{m\binom{2m}{m}} = \frac{2}{3}\pi\sqrt{3} = \nu = 3.627598728$$

$$\sum \frac{3^m}{\binom{2m}{m}} = 2\nu + 3$$

$$\sum \frac{m3^m}{\binom{2m}{m}} = 10\nu + 18$$

$$\sum \frac{m^2 3^m}{\binom{2m}{m}} = 2(43\nu + 78)$$

$$\sum \frac{(-1)^{m-1}2^m}{m\binom{2m}{m}} = \rho/3, \qquad \text{where } \rho = \sqrt{3}\log(2 + \sqrt{3}) = 2.281037989$$

$$\sum \frac{(-1)^{m-1}2^m}{\binom{2m}{m}} = \frac{\rho + 3}{9}$$

$$\sum \frac{(-1)^{m-1}m2^m}{\binom{2m}{m}} = \frac{1}{3}$$

$$\sum \frac{(-1)^{m-1} m^2 2^m}{\binom{2m}{m}} = \frac{1}{27}(3 - \rho)$$

$$\sum \frac{(-1)^m m^3 2^m}{\binom{2m}{m}} = \frac{1}{81}[\rho + 15]$$

$$\sum \frac{(2-\sqrt{2})^m}{\binom{2m}{m}} = \frac{3-2\sqrt{2}}{4}(\pi\sqrt{2} + 4)$$

$$\sum \frac{(-1)^{m-1} 3^{2m}}{4^m \binom{2m}{m}} = \frac{48}{125}\left(\log 2 + \frac{15}{16}\right)$$

$$\sum \frac{2^m (2-\sqrt{3})^m}{m^2 \binom{2m}{m}} = \frac{\pi^2}{36} = (\zeta(2))^2.$$

## References


1. Louis Comtet, Advanced Combinatorics, Dreidel, Dordrecht, 1974, p. 89.
2. H. W. Gould, Bibliography of the Euler-Fuss-Segner-Catalan Sequence, manuscript containing 243 titles.
3. Z. A. Melzak, Companion to Concrete Mathematics, vol. 1, Wiley, 1973, p. 108.
4. John Riordan, Combinatorial Identities, Wiley, 1968, pp. 131–135.
5. A. J. van der Poorten, A proof that Euler missed, Math. Intelligencer, vol. 1 (1979) 195–203.
6. ______, Some wonderful formulas, Queen's Papers in Pure and Applied Math., 54 (1979) 269–286.


# THE LOGIC OF GRAPH-THEORETIC DUALITY


T. A. McKEE

*Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435*


Dualities and duality principles are prized wherever they occur in mathematics. Some optimists see them as mechanically doubling the number of results in a theory. Others (also optimists) see them as halving the number of results, but packing more substance into each.

There are limitations which keep duality from being as powerful in graph theory as it is in many other areas, but it is still a major unifying theme. Robin Wilson's introductory text [**24**] is organized around duality, and it is also central to many area of applications. As evidence of the latter, consider the emphasis on duality in books such as Johnson & Johnson's *Graph Theory with Engineering Applications* [**8**], Price's operations research monograph *Graphs and Networks* [**19**], and Nakanishi's *Graph Theory and the Feynman Integral* [**16**].

Classical graph-theoretic duality centers on the relationship between circuits and cutsets, with spanning trees also playing a basic role. The first section of this paper surveys this duality and its formulation as a syntactical principle. A key feature of this duality is the role played by a

---

*Terry A. McKee*: After receiving a Ph.D. in mathematical logic from the University of Wisconsin-Madison in 1974, I worked with applications of logic to subjects such as topology and geometry. I have since grown increasingly fond of graph theory and now inhabit a niche which may be described as "graph metatheory." My other interests include lap swimming and studying the methodologies evidenced in vintage detective fiction.

distinguished edge. When this localization is removed, something rather surprising occurs: A second duality emerges, unexpectedly centered on the relationship between spanning trees and cutsets, yet having the same structural features as the traditional circuit/cutset duality. The second section below introduces this revisionist global duality. The third section applies this new duality to a portion of the theory of spanning trees.

**1. The Classical Circuit / Cutset Duality.** Although our graph-theoretic notation and terminology follow Wilson's text [24], it is useful to illustrate the basic concepts using examples. Review of the classical duality is especially important since part of the significance of the duality of Section 2 is its relationship to classical duality.

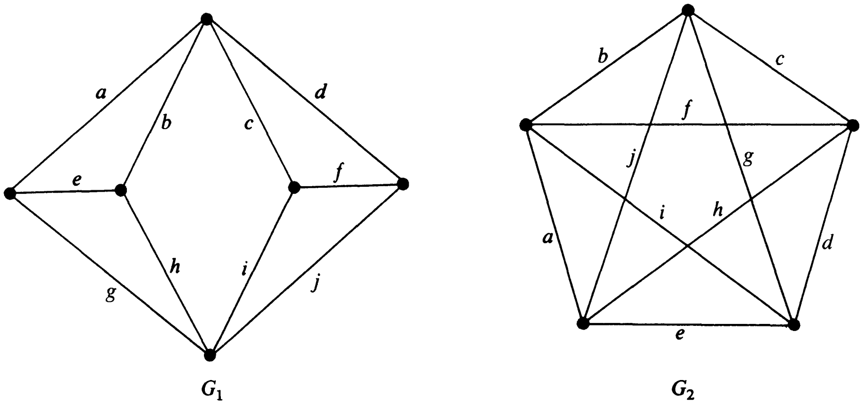



FIG. 1

Graph $G_1$ in Fig. 1 has six vertices and ten edges and is a *plane graph* in that it has been drawn without edges crossing (except, of course, at the endpoints of the edges). Graph $G_2$ has five vertices and ten edges and is *nonplanar* in that it can be shown that there is no way to embed five points in the plane and connect them with ten curves in this way without at least one crossing. Two edges are called *adjacent* whenever they share a common endpoint, and a *chain* of edges is any sequence of edges (such as $a, b, j, e$ in $G_2$) with each interior edge of the sequence adjacent to its predecessor and successor at its two endpoints. For simplicity, we consider only graphs which are finite and connected (i.e., every two vertices are joined by a chain of edges).

It is fairly common to treat vertices as the fundamental objects of graph theory, with edges merely being unordered pairs of vertices. But for studying duality (and in many applications), edges are taken as the fundamental objects. We are primarily interested in special sets of edges, calling the cardinality of such a set its *size*. A natural example is a *circuit* (i.e., a set of edges forming a chain beginning and ending at the same vertex without that vertex appearing again and without any other vertex being repeated). For instance, $G_1$ has twenty-two circuits of sizes 3, 4, 5 and 6, one of which is $\{a, c, i, g\}$; $G_2$ has thirty-seven of sizes 3, 4 and 5. An equally basic but less natural example is a *cutset* (i.e., a minimal set of edges whose removal would leave a disconnected graph). For instance, $G_1$ has twenty-two cutsets of sizes 3, 4, 5 and 6, one of which is $\{a, b, i, j\}$; $G_2$ has fifteen of sizes 4 and 6. (Check that the size-six circuits of $G_1$ are identical with its size-six cutsets.) While cutsets are fairly easily identified in plane graphs, they are harder to find in nonplanar graphs. For instance, $\{a, b, c, e, g, h\}$ is a cutset of $G_2$.

In studying duality, we allow some special kinds of edges which are frequently excluded from other parts of graph theory. In particular, we allow *loops* (i.e., edges joining vertices to themselves). Loops are thus circuits of size one and are precisely the edges which can never be in cutsets. We also allow *multiple edges* (i.e., edges joining the same pair of vertices); in other words, we allow circuits of size two.

The relationship between circuits and cutsets is the heart of graph-theoretic duality, but it is

very vague what duality means beyond that. (The same duality occurs in electrical network theory between current and voltage-drops, but it is equally vague there; see [**15**].) The best description may be that certain parts of graph theory resemble other parts when the roles of circuits and cutsets are reversed. For instance, the condition (commonly called nonseparability or two-connectedness) that every two edges are in a common circuit is equivalent to every two edges being in a common cutset. Similarly, the theory of eulerian graphs (in which every cutset has even size) somewhat resembles the theory of bipartite graphs (in which every circuit has even size). "High-level" discussion of duality are contained in [**21**] and [**7**].

The duality of circuits and cutsets suggests other dual pairs of concepts. For instance, an *isthmus* is an edge which constitutes a cutset of size one and is the only sort of edge which can never be in a circuit. So isthmuses act as duals of loops. Spanning trees are maximal sets of edges which contain no circuit (e.g., $\{a, b, h, i, j\}$ in $G_1$) and are very important in both the theory and applications of graphs. Surprisingly, the complements of spanning trees (within the total edge set) act as their duals. These complements of spanning trees are called *cotrees*.

As vague as duality is in general, it becomes very concrete for plane graphs; see [**22**]. For instance, from our example $G_1$ we can define a new graph $G_1^*$ whose vertices correspond to points from each of the six "regions" of the geometric drawing of $G_1$. Edges are drawn between vertices of $G_1^*$ precisely when the corresponding regions of $G_1$ are bordered by a common edge. In this geometric setting, circuits of $G_1$ correspond to cutsets of $G_1^*$ and vice versa, loops to isthmuses, and spanning trees to cotrees. But none of this carries over to nonplanar graphs such as $G_2$.

The opposite strategy is also useful. Instead of narrowing the family of graphs to plane graphs, we can enlarge it to matroids, as was done by Whitney [**23**]. (An excellent introductory survey occurs in [**25**].) Within matroid theory, duality is very precise and powerful. Yet graphs are more general than plane graphs and less general than matroids. Graph-theoretic duality is very different from matroid duality and is inherently limited.

Despite its limitations—and because of its vagueness—it is profitable to view graph-theoretic duality syntactically, that is, in terms of how properties can be expressed in formal languages and the effect of the symbolic interchange of circuits with cutsets. Although there is no elegant duality principle as in projective geometry or boolean algebra (or matroid theory), there are limited duality principles which are interesting. They are true duality principles in that they are equivalences of statements in which circuits and cutsets have been interchanged. But their application is restricted to very specific types of statements, and a change is also required in the logical structure of the statements.

The simplest such graph-theoretic duality principle originated with George Minty and electrical network applications in the early 1960's; see [**14**] or [**6**, page 41]. Suppose we have a graph with a distinguished edge $e$ and suppose some of the edges (possibly including $e$) are somehow "marked". Abbreviating $X - \{e\}$ as $X - e$, Minty's principle states that either

$$(\exists \text{ circuit } X \text{ containing } e)(\forall x \in X - e)(x \text{ is marked})$$

or

$$(\exists \text{ cutset } X \text{ containing } e)(\forall x \in X - e)(x \text{ is not marked}),$$

*but not both*. By letting $M(x)$ abbreviate "$x$ is marked" and doing some elementary logical manipulation, this can be rephrased as the following equivalence, holding for every graph and every distinguished edge $e$:

$$(1.1)\qquad \begin{aligned}&(\exists \text{ circuit } X \text{ containing } e)(\forall x \in X - e)\, M(x)\\ \Leftrightarrow\ &(\forall \text{ cutset } X \text{ containing } e)(\exists x \in X - e)\, M(x).\end{aligned}$$

In this form, $M(x)$ can either be thought of as identifying an arbitrary set of edges or as being an edge predicate and so any symbolic expression involving $x$ (but not $X$).

There are many ways to modify this principle and many ramifications of the modifications; these are surveyed in [**13**]. Notice that the equivalence (1.1) involves the logical duality of the

universal and existential quantifiers, as well as circuit/cutset duality. In fact [**12**] views this as identifying the quantifier duality with matrodial duality. (Edmonds and Fulkerson [**3**] also allude to this.)

To see this in more detail, notice that the formulation

$$\neg(\exists \text{ circuit } X \text{ containing } e)(\forall x \in X - e)\, M(x)$$

constitutes a generalized quantifier (in the sense of [**9**, pages 100-101]). This could be abbreviated as $QxM(x)$ with the quantifier $Qx$ translated as "for all $x$ (except possibly $e$) of some circuit containing $e$". Just as $\neg\forall x \neg M(x)$ (which is equivalent to $\exists x M(x)$) is the logical dual of $\forall x M(x)$, we can dualize this "circuit containing $e$" quantifier as

$$\neg(\exists \text{ circuit } X \text{ containing } e)(\forall x \in X - e)\neg M(x)$$

which is equivalent by (1.1) to

$$\neg(\forall \text{ cutset } X \text{ containing } e)(\exists x \in X - e)\neg M(x),$$

and so is equivalent to

$$(\exists \text{ cutset } X \text{ containing } e)(\forall x \in X - e)\, M(x).$$

Therefore these "circuit containing $e$" and "cutset containing $e$" quantifiers are logical duals of each other.

The dependence on the distinguished edge $e$ is also a conspicuous feature of (1.1). This relativization to $e$ is often incorporated into the graph by treating $e$ as being so distinguished that it is not even drawn. Instead, its endpoints—call them $\varepsilon$ and $\delta$—are distinguished as "terminals" of the remaining graph. For instance, $G_2$ could be reconfigured as shown in Fig. 2.

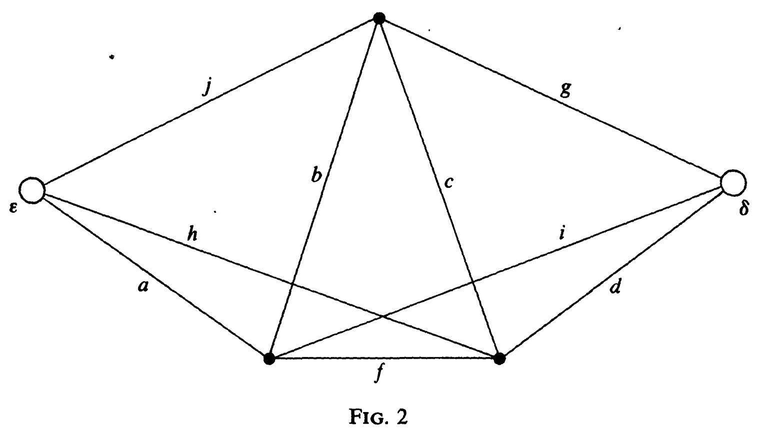



FIG. 2

The circuits containing $e$ now appear as *$\varepsilon,\delta$-chains*; i.e., as chains connecting $\varepsilon$ with $\delta$. The cutsets containing $e$ correspond to *$\varepsilon,\delta$-cutsets*; i.e., to cutsets separating $\varepsilon$ from $\delta$. This two-terminal approach can be built into (1.1) as

$$(1.2) \qquad (\forall \varepsilon,\delta)[(\exists \varepsilon,\delta\text{-chain } X)(\forall x \in X)\, M(x) \Leftrightarrow (\forall \varepsilon,\delta\text{-cutset } X)(\exists x \in X)\, M(x)].$$

As one advantage of this formulation, it can be easily related to the max-flow min-cut theorem and other parts of network flow theory. It also emphasizes the very fundamental dual-like relationship between connection and separation.

**2. The Tree / Cutset Duality.** What makes equivalence (1.2) local is the dependence on the vertex pair $\varepsilon,\delta$. This can be avoided by "distributing" the $(\forall \varepsilon,\delta)$ quantifier across the $\Leftrightarrow$ symbol:

$$(\forall \varepsilon,\delta)(\exists \varepsilon,\delta\text{-chain } X)(\forall x \in X)\, M(x) \Leftrightarrow (\forall \varepsilon,\delta)(\forall \varepsilon,\delta\text{-cutset } X)(\exists x \in X)\, M(x).$$

The right side talks about all cutsets of the graph; $\varepsilon$ and $\delta$ are irrelevant. Letting *tree* hereafter abbreviate "spanning tree," the left side is easily rephrased as

$$(\exists \text{ tree } X)(\forall x \in X)\, M(x).$$

Thus this equivalence is naturally expressible as the following global tree/cutset duality principle:

$$(2.1) \qquad (\exists \text{ tree } X)(\forall x \in X)\, M(x) \Leftrightarrow (\forall \text{ cutset } X)(\exists x \in X)\, M(x).$$

There is an alternate form of (2.1) which is essentially its contrapositive:

$$(\forall \text{ tree } X)(\exists x \in X)\, M(x) \Leftrightarrow (\exists \text{ cutset } X)(\forall x \in X)\, M(x).$$

Yet there is also a true dual in the sense of classical duality:

$$(2.2) \qquad (\exists \text{ cotree } X)(\forall x \in X)\, M(x) \Leftrightarrow (\forall \text{ circuit } X)(\exists x \in X)\, M(x).$$

This occurs automatically because (2.1) holds for all matroids (translating "tree" as "base" and "cutset" as "cocircuit"). Thus there is a global cotree/circuit duality as well.

Although it is rather jarring to suddenly start thinking of spanning trees as being duals of cutsets, there is certainly some sort of intimate relationship involved. A clear intimation of this occurs in Carré [2, page 160]:

> It is evident that a spanning tree represents a minimal collection of edges which preserves the connectedness of a graph. This concept is in a sense complementary to that of a proper cut set of edges (which is a minimal collection of edges whose removal disconnects some nodes from others). These notions are related precisely by the following theorem:
>
> *In a connected graph, every cut set of edges has at least one edge in common with every spanning tree.*

The theorem he cites corresponds to the left-to-right direction of equivalence (2.1) by letting $M(x)$ be the membership of $x$ in any particular spanning tree.

From more abstract viewpoints, it is very natural to call this tree/cutset relationship a duality. For instance, it is Example 4 of [4], illustrating clutter/blocker duality and so inheriting some structure such as Fulkerson's Max-Min Theorem [4, Theorem 2.1]. It is also an example of Woodall's abstract notion of Menger duality [**26**]. But what role can this tree/cutset duality play in graph theory itself, and how is it related to classical circuit/cutset duality?

Consider the natural quantifier corresponding to the family of all spanning trees of a graph:

$$(\exists \text{ tree } X)(\forall x \in X) \ldots.$$

By exactly the same method as we used in Section 1 to dualize the "circuit containing $e$" quantifier, except using (2.1) instead (1.1), we see that this tree quantifier is the logical dual of the cutset quantifier

$$(\exists \text{ cutset } X)(\forall x \in X) \ldots.$$

Similarly, the natural cotree and circuit quantifiers are duals of each other.

We have thus come full circle. We saw that the classical circuit/cutset duality corresponds precisely to logical duality as long as everything is relativized to a distinguished edge $e$. In particular, (1.1) is a localized duality principle. When the localization is dropped from, say, the "cutset containing $e$" quantifier, spanning trees suddenly emerge as the duals in the global sense. Moreover (2.1), the fundamental equivalence underlying tree/cutset duality, is precisely the global form of the fundamental equivalence of circuit/cutset duality. This is summarized in Fig. 3.

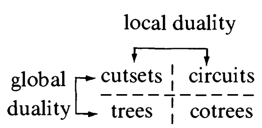



FIG. 3

Since spanning trees are sometimes called "minimal connectors", call a set of edges a *connecting set* if it contains a spanning tree of the graph. Call a set a *separating set* if it contains a cutset. Then equivalence (2.1) also holds with "tree" replaced by "connecting set" and "cutset" by "separating set". Thus connecting and separating become duals, and starting with cotrees and circuits leads to a similar duality between dependence (i.e., containing a circuit) and coindependence.

Each cutset has at least one edge in common with each spanning tree. It is also important that, given any spanning tree $T$ and any edge $e$ in $T$, there is a cutset which contains $e$ but is otherwise disjoint from $T$. This can be derived from a statement about trees which Whitney [**23**] took as a fundamental property:

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)(\forall x \in X - Y)(\exists y \in Y - X)(X - x + y \text{ is a tree}),$$

where $X - x + y$ abbreviates $(X - \{x\}) \cup \{y\}$. Rephrasing this as

$$(\forall \text{ tree } X)(\forall x \in X)(\forall \text{ tree } Y)(\exists y \in Y)\Big[y = x \text{ or}\Big(y \notin X \text{ and } X - x + y \text{ is a tree}\Big)\Big]$$

allows the application of (2.1), producing

$$(\forall \text{ tree } X)(\forall x \in X)(\exists \text{ cutset } Y)(\forall y \in Y)\Big[y = x \text{ or}\Big(y \notin X \text{ and } X - x + y \text{ is a tree}\Big)\Big].$$

Taking $X = T$ and $x = e$ produces the desired cutset $D = Y$, and there still is a bit left over from Whitney's formulation. If the graph has $n$ vertices, each spanning tree has $n - 1$ edges, so each edge $y$ of $D$ produces a different spanning tree containing the $n - 2$ edges of $T - e$. Moreover, each such spanning tree must meet $D$ for its final edge, and so each determines a different edge from $D$. Thus Whitney's property plus (2.1) corresponds to the following theorem of Hakimi [**5**], which is particularly interesting since it states a tree/cutset relationship somewhat resembling the classical max-flow min-cut theorem.

PROPOSITION. *Suppose $G$ is any graph having $n$ vertices. Every cutset of $G$ contains at least $k$ edges if and only if every set of $n - 2$ edges from any spanning tree of $G$ is contained in at least $k$ spanning trees.*

**3. An Illustration: Maximal Distance Between Trees.** The family of all spanning trees of a graph received a good deal of attention during the 1960's in the literatures of both graph theory and electrical engineering; see [**1**, Sect. 164]. A key idea is a notion introduced by Ore [**18**, page 104] of distance between two spanning trees: namely, the number of edges in either but not the other. Equivalently, each connected graph $G$ determines a *tree graph* $T(G)$ having vertices which correspond to spanning trees of $G$. Two vertices of $T(G)$ are joined by an edge precisely when the corresponding spanning trees agree on all but one interchanged pair of edges; i.e., when they are a distance one apart. Tree graphs enjoy many pleasant properties (e.g., they are always hamiltonian) and reflect a great deal of information about the original graph (e.g., they determine the original up to circuit isomorphism).

Our approach yields a self-dual circuit/cutset formulation of the maximal distance between spanning trees of a graph or, equivalently, the diameter of its tree graph.

THEOREM. *The maximal distance between spanning trees of a connected graph $G$ is less than $k$ if and only if every set of $k$ edges of $G$ contains either a circuit or a cutset (i.e., is either dependent or separating).*

*Proof.* The property

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)(|X - Y| < k)$$

can be re-expressed in the following logically equivalent reformulations, each preceded by a

universal quantification over all sets $S$ having $k$ edges:

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)[S \subseteq X \to S \cap Y \neq \varnothing],$$

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)\Big[\big(\forall x \notin X\big)\big(x \notin S\big) \to (\exists y \in Y)(y \in S)\Big],$$

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)\Big[\big(\exists x \notin X\big)(x \in S) \quad \text{or} \quad (\exists y \in Y)(y \in S)\Big],$$

$$(\forall \text{ tree } X)\big(\exists x \notin X\big)(x \in S) \quad \text{or} \quad (\forall \text{ tree } Y)(\exists y \in Y)(y \in S),$$

$$(\forall \text{ cotree } X)(\exists x \in X)(x \in S) \quad \text{or} \quad (\forall \text{ tree } Y)(\exists y \in Y)(y \in S).$$

Using equivalences (2.1) and (2.2) on the last, we get for every set $S$ of size $k$

$$(\exists \text{ circuit } X)(\forall x \in X)(x \in S) \quad \text{or} \quad (\exists \text{ cutset } Y)(\forall y \in Y)(y \in S),$$

thereby proving the theorem.

The $k = 1$ case is trivially true, since the maximal distance between spanning trees is zero if and only if the graph is (except possibly for loops) itself a tree, and this is clearly equivalent to every edge either being a circuit (i.e., a loop) or a cutset (i.e., an isthmus). If $n$ is the number of vertices of $G$, the $k = n$ case is also trivial: Every spanning tree is a maximal circuitless subgraph and has $n - 1$ edges, so every set of $n$ edges contains a circuit. (Dually, if $q$ is the number of edges of $G$, the $k = q - n + 2$ case is trivial, since every set of $q - n + 2$ edges contains a cutset.)

But the cases adjacent to these trivial ones are interesting. When $k = 2$, the theorem characterizes every two distinct spanning trees being distance one apart; i.e., the tree graph being complete (see [**10**]). This happens if and only if each pair of edges contains either a circuit or a cutset. If there are no loops, these are trees and the unicyclic graphs.

The $k = n - 1$ case gives the following corollary. (Other approaches to pairwise intersection of spanning trees occur in [**20**] and [**17**], and in [**11**].)

COROLLARY. *Every two trees intersect if and only if every set of $n - 1$ edges contains either a circuit or a cutset.*

This corollary can also be reached directly, since the pairwise intersection of trees can be expressed as

$$(\forall \text{ tree } X)(\forall \text{ tree } Y)(\exists y \in Y)(y \in X) \tag{3.1}$$

and so by (2.1) is equivalent to

$$(\forall \text{ tree } X)(\exists \text{ cutset } Y)(\forall y \in Y)(y \in X). \tag{3.2}$$

The latter simply says that every spanning tree contains a cutset, which is essentially the content of the corollary. If a set does not contain a spanning tree, its complement must contain a cutset, so (3.2) is also equivalent to "every set either contains a cutset or its complement contains a cutset".

It is intriguing that the tree/cutset dual of the above—the pairwise intersection of cutsets—is equivalent to the graph being complete (i.e., every two vertices are joined by an edge) if there are no multiple edges. The discussion following the corollary has parallels for this version; i.e., "every cutset contains a tree" and "every set either contains a tree or its complement contains a tree". But there does not seem to be a parallel version of the corollary itself.

## References


**1.** W. G. Brown (Editor), Reviews in Graph Theory, Amer. Math. Soc., Providence, 1980.

**2.** B. Carré, Graphs and Networks, Clarendon Press, Oxford, 1979.

**3.** J. Edmonds and D. R. Fulkerson, Bottleneck extrema, J. Combin. Theory, 8 (1970) 299–306.

**4.** D. R. Fulkerson, Networks, frames, blocking systems, in Mathematics of the Decision Sciences, Part 1, G. B. Dantzig, Editor, Amer. Math. Soc., Providence, 1968, pp. 303–334.

5. S. L. Hakimi, On the realizability of a set of trees, I. R. E. Trans. Circuit Thy, CT8 (1961) 11–17.

6. F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.

7. M. Iri, "Dualities" in graph theory and the related fields viewed from the metatheoretic standpoint, in Graph Theory and Algorithms, Tsuken Symposium, Tohoku Univ., 1980, pp. 139–151.

8. D. E. Johnson and J. R. Johnson, Graph Theory with Engineering Applications, Ronald Press, New York, 1972.

9. H. J. Keisler, Fundamentals of model theory, in Handbook of Mathematical Logic, J. Barwise, Editor, North-Holland, Amsterdam, 1977, pp. 47–103.

10. G. Kishi and Y. Kajitani, Generation of Hamilton circuits in tree graphs, Electron. Commun. Japan, 50 (1967) 40–47.

11. S. Kundu, Bounds on the number of disjoint spanning trees, J. Combin. Theory Ser. B, 17 (1974) 199–203.

12. T. A. McKee, A quantifier for matroid duality, Discrete Math., 34 (1981) 315–318.

13. ______, Recharacterizing eulerian: Intimations of new duality, Discrete Math., 51 (1984) 237–242.

14. G. J. Minty, On the axiomatic foundations of the theories of directed linear graphs, electrical networks and network programming, J. Math. Mech., 15 (1966) 485–520.

15. ______, On the duality-principle of electrical network theory, in Théorie des Graphes, Dunod, Paris, 1967, pp. 249–252.

16. N. Nakanishi, Graph Theory and Feynman Integrals, Gordon and Breach, New York, 1971.

17. C. St. J. A. Nash-Williams, Edge-disjoint spanning trees of finite graphs, J. London Math. Soc., 36 (1961) 445–450.

18. O. Ore, Theory of Graphs, American Mathematical Society Colloquium Publications, Vol. 38, Amer. Math. Soc., Providence, 1962.

19. W. L. Price, Graphs and Networks, Butterworth, London, 1971.

20. W. T. Tutte, On the problem of decomposing a graph into $n$ connected factors, J. London Math. Soc., 36 (1961) 221–230.

21. L. Weinberg, Duality in networks, Networks, 5 (1975) 179–201.

22. H. Whitney, Non-separable and planar graphs, Trans. Amer. Math. Soc., 34 (1932) 339–362.

23. ______, On the abstract properties of linear dependence, Amer. J. Math., 57 (1935) 509–533.

24. R. J. Wilson, Introduction to Graph Theory, Oliver & Boyd, London, 1972.

25. ______, An introduction to matroid theory, this Monthly, 80 (1973) 500–525.

26. D. R. Woodall, Menger and Konig systems, in Theory and Applications of Graphs, Y. Alavi & D. R. Lick, Editors, Lecture Notes in Mathematics, Vol. 642, Springer-Verlag, Berlin, 1978, pp. 620–635.

---

## 151. MISCELLANEA

An old mathematician can grow indefinitely, what he cannot do is keep up. Cultural conservatism is becoming in an older mathematician: anything else is cosmetics anyway. If he whores after the new thing, he will only get it wrong and wind up praising the latest charlatans.

—Wilfrid Sheed (adapted by R. P. Boas).

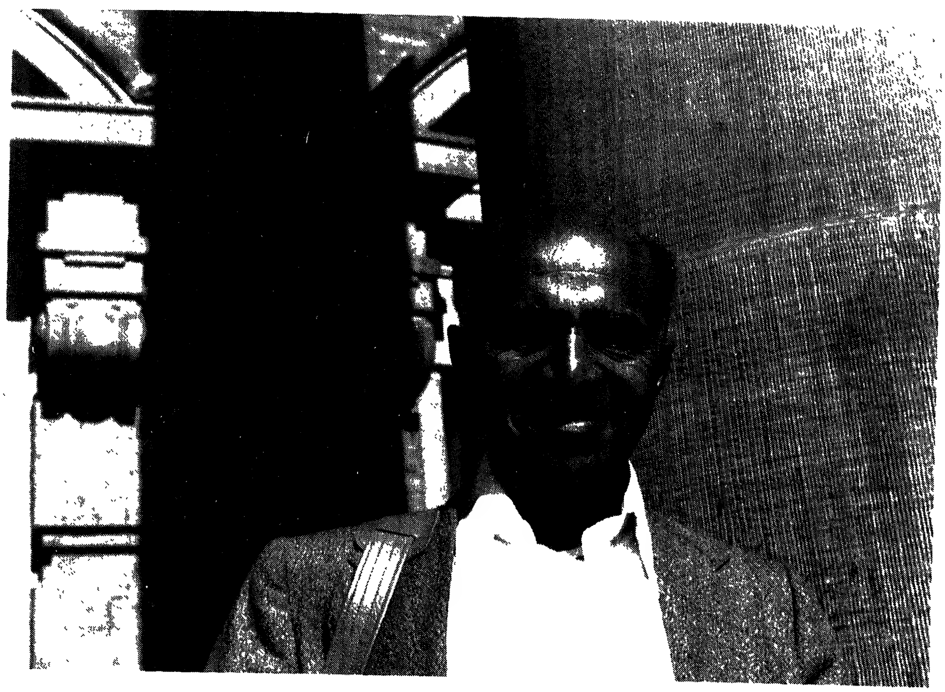
(The original had "writer" instead of "mathematician".)

---

## 152. MISCELLANEA

There are more than twice as many left-handed architects and mathematicians as one would predict from the general population.

—From a review by Michael S. Gazzaniga of *Cerebral Dominance* by Norman Geschwind and Albert M. Galaburda (Harvard University Press, Cambridge, MA, 1984).



Is the mathematics that this man does honest statistics or honest measure theory? (See p. 490.)

$$\begin{aligned}(1-yx)w &= (1-yx)(1+yzx)\\ &= 1-yx+yzx-yxyzx\\ &= 1-yx+y[(1-xy)z]x,\end{aligned}$$

which is 1, because the quantity in brackets is 1.

A closer look at this computation shows something else: we have only used the assumption that $z$ is a *right* inverse of $1-xy$ and have deduced that $w$ is then a *right* inverse of $1-yx$. The same is true with left in place of right, because

$$(1+yzx)(1-yx) = 1-yx+y[z(1-xy)]x.$$

Thus we get the following result, which actually does more than just answer our two questions:

THEOREM 1. *If $z$ is a right [left] inverse of $1-xy$, then $1+yzx$ is a right [left] inverse of $1-yx$.*

As we just saw, the proof of this theorem is a total triviality. But if we had been unwilling to use infinite series in a context where they may make no sense, how difficult would it have been to *discover* $1+yzx$?

When $y=1$, our two questions are of course pointless, but Theorem 1 tells us something even then that we might not have noticed otherwise, namely:

*If $z$ is a right inverse of $1-x$, so is $1+zx$.*

Let us now assume that $1-x$ has a unique right inverse $z$. Then it follows that $1+zx=z$. This implies $z(1-x)=1$, *so that $z$ is also a left inverse of $1-x$!* In other words, $1-x$ is invertible. Since every element of $R$ can be written in the form $1-x$, we have arrived at the result to which the title of this note alludes.

THEOREM 2. *The invertible elements of $R$ are precisely those that have unique right inverses in $R$.*

Of course, the same is true with left in place of right.

The following well-known fact from linear algebra is also an immediate consequence of Theorem 1:

*If $A$ and $B$ are $n$-by-$n$ matrices over some field, then $AB$ and $BA$ have the same eigenvalues.*

P.S. After completing this note, I was told that the geometric series trick of finding $1+yzx$ is described in [1] and [2]. However, no conclusions about one-sided inverses are drawn there. In the context of Banach algebras (where series do make sense) $1+yzx$ occurs in an exercise on p. 259 of [4]. The referee has pointed out that Theorem 2 appears as Exercise 6 on p. 89 of [3].

### References

1. P. R. Halmos, Does mathematics have elements? The Math. Intelligencer, 3 (1981) 147–153.
2. ______, Does mathematics have elements? Bull. Austral. Math. Soc., 25 (1982) 161–175.
3. N. Jacobson, Basic Algebra, vol. 1, Freeman, 1974.
4. W. Rudin, Functional Analysis, McGraw-Hill, 1973.

---

**ANSWER TO PHOTO ON PAGE 465**

David Blackwell.

# INEQUALITIES ABOUT SYMMETRIC COMPACT CONVEX SETS IN THE PLANE

MARC ROGALSKI
*Université de Lille I, U.E.R. de Mathématiques Pures et Appliquées, 59655-Villeneuve D'Ascq Cedex, France*

JEAN SAINT RAYMOND
*Université Pierre et Marie Curie, Place Jussieu, 75005-Paris, France*

**Dedicated to Professor G. Choquet on the occasion of his seventieth birthday**

Let $\mathscr{C}$ be a symmetric compact convex set in the euclidean plane, $D$ its diameter, $L$ its perimeter, $\Sigma$ its area. The circle $\Gamma$ of radius $D/2$, with the same center as $\mathscr{C}$, surrounds $\mathscr{C}$ and touches it at antipodal points $A$ and $B$ (see Fig. 1). Hence, it is obvious that $\Sigma \leqslant \pi D^2/4$ and $2D \leqslant L \leqslant \pi D$. Are there other inequalities among these three numbers? Yes, of course, and many of them were found many years ago. They correspond generally to figures solving an extremal problem; for instance, if $\Sigma$ and $D$ are given, what is the maximum of $L$? (Favard, 1929, [**2**]). Or, if $\Sigma$ is given, what is the minimum of $L$? (the isoperimetric inequality: $L^2 \geqslant 4\pi\Sigma$, proved by Weierstrass).

About a new inequality [**5**], we shall compare old and new methods, geometric and analytic points of view, and give elementary partial solutions, useful for teaching aims, or more sophisticated ones.

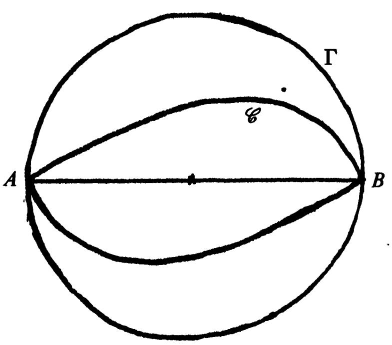



FIG. 1

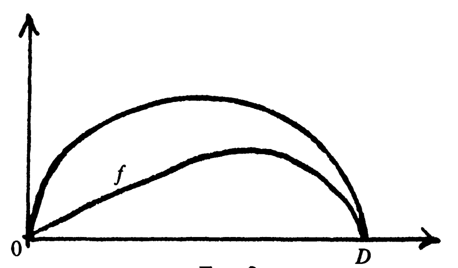



FIG. 2

An elementary but useful remark is that, if we know the diameter $D$ of $\mathscr{C}$, then $\mathscr{C}$ is defined by a concave continuous function $f$ on the interval $[0, D]$, satisfying the inequality $f(x) \leqslant \sqrt{x(D-x)}$ (see Fig. 2). In such a situation, we denote by $l$ the length of the graph of $f$, and by $S$ its area, defined by $S = \int_0^D f(x)\,dx$; obviously $L = 2l$ and $\Sigma = 2S$. With this point of view, geometric problems on figures become analytic problems on functions.

---

*Marc Rogalski*: I was born in Paris, France, in 1940. I received my "Thèse d'Etat" in 1972 from the University of Paris 6, under the supervision of Professor G. Choquet, for works in compact convex sets theory and ordered Banach spaces. I have taught first at the University of Orsay in 1972, then at the University of Limoges during a year, and now at the University of Lille. My interest in mathematics is at present in geometry of Banach spaces. My chief extramathematical interests are in syndicalism, politics, and yachting.

J. Saint Raymond was born in 1946. He studied mathematics at the Ecole Normale Supérieure from 1965 to 1969, and obtained his "Thèse d'Etat" in 1977 under the direction of G. Choquet. Since 1980 he is professor at the University of Paris 6. He works especially on descriptive theory of sets and functional analysis. Beside mathematics, he likes playing music.

The literature about convex sets is very extensive. We refer the reader specially to **[1]** and **[3]**, and to their bibliography.

**1. Old inequalities.** We start with two inequalities which are useful as a guide toward our problem.

(a) *Favard's first inequality* (1929, **[2]**).

By solving geometrically the problem to determine the convex sets (not necessarily symmetric) of maximal perimeter $L$ for $\Sigma$ and $D$ given, Favard proved the following inequality:

$$LD - 2\Sigma \leqslant 2D^2. \tag{1}$$

(b) *Favard's second inequality* (1929).

In the same paper, he proved for a concave continuous function on $[0, D]$, null at 0 and $D$, the following inequality

$$lD - 2S \leqslant \sqrt{D^4 + 4S^2} \tag{2}$$

(in the case where $D$ is the diameter of the symmetric convex set $\mathscr{C}$ defined by $f$, i.e., $f(x) \leqslant \sqrt{x(D-x)}$, inequality (1) implies inequality (2)). Because the maximum of $f$ on $[0, D]$ is less than $l/2$, it is obvious that

$$0 \leqslant lD - 2S. \tag{3}$$

Favard determines the function $f$ that maximizes $l$ for given $S$. His proof is geometric in nature, and easy analytic arguments are omitted. Later we shall give a modern complete proof of Favard's inequalities, in the symmetric case.

**2. A new problem.** Inequality (1) is equivalent to

$$\Sigma \geqslant D\left(\frac{L}{2} - D\right), \tag{4}$$

and from (2) and (3) we derive

$$\Sigma \geqslant \frac{D}{L}\left(\frac{L}{2} + D\right)\left(\frac{L}{2} - D\right),$$

which is less good.

This suggests a question. Obviously, if $\mathscr{C}$ is a segment, $L/2 = D$ and $\Sigma = 0$. In the other case, $\frac{L}{2} - D > 0$ and $\Sigma > 0$, and *it is reasonable to think that there exists a constant* $C > 0$ *such that*

$$\Sigma \geqslant C\left(\frac{L}{2} - D\right)^2 \tag{5}$$

(the square is for homogeneity). This question was put to us by A. Pajor, and its positive answer is a particular case of a conjecture about mixed volumes and means of volumes in any dimension.

*Actually the answer is yes.* From inequality (4), we could obtain (5) if $C\left(\frac{L}{2} - D\right) \leqslant D$, or $C \leqslant \frac{x}{\frac{1}{2} - x}$ for $x = \frac{D}{L}$; but the function $x \rightsquigarrow \frac{x}{\frac{1}{2} - x}$ is increasing on $\left[0, \frac{1}{2}\right[$, and $\frac{1}{2} \geqslant \frac{D}{L} \geqslant \frac{1}{\pi}$; hence, on $\left[\frac{1}{\pi}, \frac{1}{2}\right[$ ,

$$\frac{x}{\frac{1}{2} - x} \geqslant \frac{1}{\pi\left(\frac{1}{2} - \frac{1}{\pi}\right)},$$

and the inequality (5) is true for the value $C_1$ of $C$ given by

(6) $$C_1 = \frac{2}{\pi - 2} \sim 1.7519\ldots.$$

But the proof of inequality (4) that is given in [2] is not elementary. Hence *we shall give first an elementary proof of the existence of C*. Then, *our problem will be to find the greatest constant C that satisfies inequality* (5). We shall study these problems only in the case of symmetric convex sets.

REMARK 1. The quantity $\dfrac{\Sigma}{\left(\frac{L}{2} - D\right)^2}$ is not bounded from above, even for polygons which are inscribed in the circle of diameter $D$: for a rectangle of diagonal $D$ and side $a$, $\dfrac{\Sigma}{\left(\frac{L}{2} - D\right)^2} \sim \dfrac{D}{a}$ when $a \to 0$.

### 3. An elementary proof of the existence of the constant $C$

PROPOSITION 1. *For a symmetric convex compact set* $\mathscr{C}$,

$$\Sigma \geqslant C_2\left(\frac{L}{2} - D\right)^2, \text{ where } C_2 = \inf_{0<\theta\leqslant\pi/2} \frac{\sin\theta}{2(\theta - 1 + \cos\theta)^2} \geqslant 1.436.$$

*Proof*. (a) We want to prove the inequality $S \geqslant C_2\dfrac{(l-D)^2}{2}$, for $f$ concave and continuous on $[0, D]$, satisfying $f(x) \leqslant \sqrt{x(D-x)}$. Let $m = \dfrac{D}{2}\sin\theta$ be the maximum of $f$. We have

$$S \geqslant \frac{Dm}{2} = \frac{D^2}{4}\sin\theta,$$

and

$$l \leqslant 2\left(\frac{D}{2}\theta + \frac{D}{2}\cos\theta\right) = D(\theta + \cos\theta) \text{ (see Fig. 3)};$$


FIG. 3

hence

$$\frac{(l-D)^2}{2} \leqslant \frac{D^2}{2}(\theta - 1 + \cos\theta)^2.$$

We shall have $S \geqslant \dfrac{C(l-a)^2}{2}$ if

$$\frac{D^2}{4}\sin\theta \geqslant C\frac{D^2}{2}(\theta - 1 + \cos\theta)^2.$$

The greatest $C$ compatible with this condition is

$$C_2 = \operatorname*{Inf}_{0<\theta\leqslant\pi/2} \frac{\sin\theta}{2(\theta - 1 + \cos\theta)^2},$$

and this number is strictly positive.

(b) It is easy to see that $C_2 \geqslant 1$: the function

$$u(\theta) = \sin\theta - 2(\theta - 1 + \cos\theta)^2$$

satisfies $u(0) = 0$ and $u' > 0$ on $\left[0, \frac{\pi}{2}\right]$; hence $u \geqslant 0$, and $C_2 \geqslant 1$.

But if we define $g$ by

$$g(\theta) = \frac{\sin\theta}{2(\theta - 1 + \cos\theta)^2},$$

$h$ by

$$h(\theta) = 2g'(\theta)\frac{(\theta - 1 + \cos\theta)^3}{\cos\theta},$$

and $\Psi$ by

$$\Psi(\theta) = \cos^2\theta h'(\theta),$$

we have $\Psi'(\theta) = \cos\theta(-3x^2 - 2x + 3)$, with $x \doteq \sin\theta \in [0,1]$; the only root in $[0,1]$ is $x = (\sqrt{10} - 1)/3$. Hence, it is elementary to deduce that $g$ is decreasing from infinity to a minimum on $]0, \theta_0]$, and increasing on $\left[\theta_0, \frac{\pi}{2}\right]$, and that $\theta_0$ is in the interval $[0.935, 0.937]$; as the function $\theta \rightsquigarrow \theta - 1 + \cos\theta$ is nondecreasing and nonnegative, it is easy to find that $C_2 \geqslant 1.436$.

Of course, because of its elementary nature, this proof of the existence of $C$ gives a bad value for $C_2$, lower than $C_1$.

**4. How to work with polygons?** It is often useful to work with polygons rather than with arbitrary convex sets. This is possible for two reasons. First, the perimeter of a convex set is approached by the perimeter of a polygon whose edges are short chords of the convex set (by definition). Second, the area of such a polygon is an approximation to the area of the convex set (obvious from the formula $\frac{1}{2}\int_0^{2\pi}\rho^2(\theta)\,d\theta$ for the area, in polar coordinates).

As an illustrative example, this method gives an easy proof of the fact that if $\mathscr{C} \subset \mathscr{C}'$, then $L(\mathscr{C}) \leqslant L(\mathscr{C}')$. The result is obvious if $\mathscr{C}$ and $\mathscr{C}'$ are polygons; otherwise there exist polygons $P \subset \mathscr{C}$ and $P' \subset \mathscr{C}'(1 + \varepsilon)$ (we suppose $0 \in \mathscr{C}$), with $P' \supset P$, and $(1 - \varepsilon)L(\mathscr{C}) \leqslant L(P) \leqslant L(P') \leqslant (1 + \varepsilon)L(\mathscr{C}')$. From this result, we deduce another which will be useful:

LEMMA 1. *On the set of concave (or convex) continuous functions on $[a, b]$, the function $f \rightsquigarrow l(f)$, where $l(f)$ is the length of the graph of $f$, is 4-lipschitzian for the uniform norm, hence continuous.*

If $f - \varepsilon \leqslant g \leqslant f + \varepsilon$, and $\alpha \leqslant \operatorname{Inf}_x f(x) - \varepsilon$, we apply the preceding result to the convex delimited by the segment $y = \alpha$, the lines $x = a$ and $x = b$, and, respectively, the graph of $f - \varepsilon$, $g$ and $f + \varepsilon$; we obtain then $|l(f) - l(g)| \leqslant 4\varepsilon$ (see Fig. 4).

Thus, if for instance the diameter $D$ of a symmetric convex set is given, any inequality depending continuously on $\Sigma$ and $L$ is true if it is true for symmetric polygons. As an application of this remark, we give a result which will be useful for an analytical proof of the Favard's first inequality, and for the determination of the greatest constant $C$ satisfying the inequality (5).

PROPOSITION 2. *Let $\mathscr{E}$ be the convex set of concave continuous functions $f$ on $[0, D]$ satisfying $0 \leqslant f(x) \leqslant \sqrt{x(D - x)}$, and let $\varphi$: $\mathscr{E} \to R$ be a function that is continuous (for the uniform norm)*

*and concave. If* $\varphi(f) \geqslant a$ *when* $f$ *is a polygon with its vertices on the circle* $y = \sqrt{x(D-x)}$, *then* $\varphi(f) \geqslant a$ *for every* $f$ *in* $\mathscr{E}$.

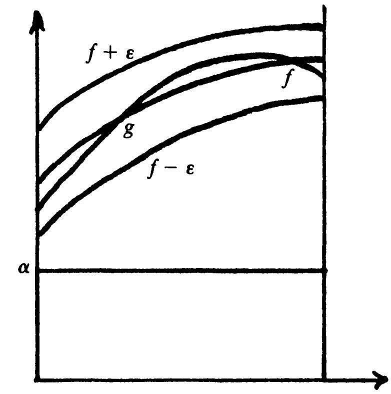
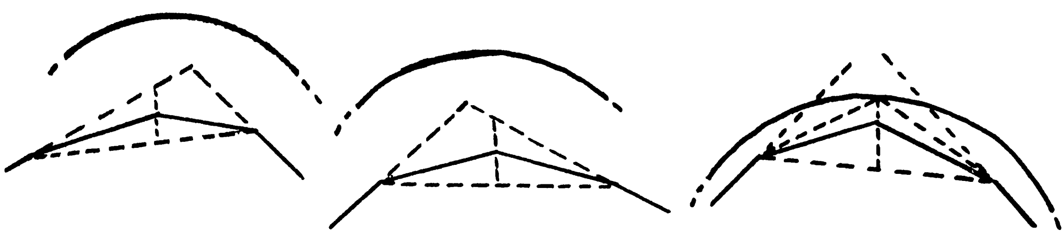



FIG. 4

To prove that $\varphi(f) \geqslant a$ for every $f$ in $\mathscr{E}$, it is sufficient by Lemma 1 to prove this inequality when the graph of $f$ is a polygon; for a vertex of such a polygon, we shall talk of "a vertex of $f$". We shall prove that if $f$ has not got all its vertices on the circle, then there exists a polygon $g$ in $\mathscr{E}$ such that $\varphi(g) \leqslant \varphi(f)$, and such that either $g$ has fewer vertices than $f$, or has the same vertices as $f$, except for one, and this one is on the circle for $g$ but not for $f$.

If a vertex $m$ of $f$ is not on the circle, three figures are possible (if the nearest vertices $p$ and $q$ of $m$ are on the circle, only the case (c) is realized). Let $\tilde{f}$ be the function whose graph is a polygon with all the vertices of $f$ except $m$ (see Fig. 5). Then there exists a continuous piecewise affine function $u$ and $t_0 \in ]0,1[$ such that $f = \tilde{f} + t_0 u$. For $t \in [0,1]$, let $g_t = \tilde{f} + tu$. For a good (and clear) choice of $u$, $g_t \in \mathscr{E}$ for every $t \in [0,1]$. Because the function $\Psi(t) = \varphi(g_t)$ is concave and continuous on $[0,1]$, its minimum is attained at 0 or 1. Hence, for $g = g_0$ or $g = g_1$, $\varphi(g) \leqslant \varphi(g_{t_0}) = \varphi(f)$ and $g$ is the function we wish.



(a) (b) (c)

FIG. 5

By applying this procedure a finite number of times, we find $g$, which satisfies $\varphi(g) \leqslant \varphi(f)$ and which either is the null function (which has its vertices on the circle!), or a genuine polygon with its vertices on the circle; we have $\varphi(f) \geqslant \varphi(g) \geqslant a$, and this proves the proposition.

COROLLARY 1. (a) *Favard's first inequality for convex symmetric sets is true if and only if it is true for symmetric polygons inscribed in the circle of diameter D.*

(b) *A constant C satisfies* $\Sigma \geqslant C\left(\frac{L}{2} - D\right)^2$ *for every symmetric convex set if and only if it satisfies this inequality for symmetric polygons inscribed in the circle of diameter D.*

It suffices to apply Proposition 2 to the function $\varphi_1$ and $\varphi_2$ defined by

$$\varphi_1(f) = S(f) - \frac{D}{2}(l(f) - D)\left(\text{where } S(f) = \int_0^D f(x)\,dx\right),$$

and

$$\varphi_2(f) = S(f) - C\frac{[l(f) - D]^2}{2};$$

indeed, $f \rightsquigarrow S(f)$ is affine, and

$$f \rightsquigarrow l(f) = \operatorname*{Sup}_n \sum_{K=0}^{n-1}\left(\frac{D^2}{n^2} + \left|f\left(\frac{K+1}{n}D\right) - f\left(\frac{K}{n}D\right)\right|^2\right)^{1/2}$$

is convex, hence $f \rightsquigarrow l(f) - D$ is convex positive, and $(l(f) - D)^2$ is too.

COROLLARY 2. (a) *Favard's first inequality for convex symmetric sets is equivalent to the relation*

$$\sum_{i=1}^{n} \sin\theta_i - 4\left(\sum_1^n \sin\frac{\theta i}{2} - 1\right) \geqslant 0, \quad \text{for} \quad \theta_i \geqslant 0, \quad \sum_1^n \theta_i = \pi. \tag{7}$$

(b) *The greatest constant $C$ satisfying the inequality $\Sigma \geqslant C\left(\frac{L}{2} - D\right)^2$ for symmetric compact convex sets is the same as the greatest constant $C$ satisfying*

$$\sum_{i=1}^{n} \sin\theta_i - 4C\left(\sum_1^n \sin\frac{\theta i}{2} - 1\right)^2 \geqslant 0, \quad \text{for} \quad \theta_i \geqslant 0, \quad \sum_1^n \theta_i = \pi. \tag{8}$$

This is obvious, because the first members of (7) and (8) are exactly the quantities $\frac{8}{D^2}\varphi_1(f)$ and $\frac{8}{D^2}\varphi_2(f)$, respectively, for $f$ an inscribed polygon with angles at the center $\theta_1, \theta_2, \ldots, \theta_n$.

**5. An analytic proof of Favard's first inequality.** We prove two lemmas in order to establish inequality (7).

LEMMA 2. *Define $\rho(x) = 2\sin x - \frac{1}{2}\sin 2x$, and, for $0 \leqslant \alpha \leqslant \pi/2$, $\sigma(x) = \rho(x) + \rho(\alpha - x)$. Then, if $0 \leqslant x \leqslant \alpha$, $\sigma(x) \leqslant \rho(\alpha)$.*

It suffices to look at the case where $x \in \left[0, \frac{\alpha}{2}\right]$. We have $\sigma(0) = \rho(\alpha)$, and

$$\begin{aligned}\sigma'(x) = \rho'(x) - \rho'(\alpha - x) &= 2\cos x - 2\cos(\alpha - x) - [\cos 2x - \cos 2(\alpha - x)]\\ &= -4\sin\frac{\alpha}{2}\sin\left(\frac{\alpha}{2} - x\right)\left[2\cos\frac{\alpha}{2}\cos\left(\frac{\alpha}{2} - x\right) - 1\right];\end{aligned}$$

but, if $0 \leqslant x \leqslant \frac{\alpha}{2} \leqslant \frac{\pi}{4}$, then

$$2\cos\frac{\alpha}{2}\cos\left(\frac{\alpha}{2} - x\right) - 1 \geqslant 2\cos^2\frac{\alpha}{2} - 1 = \cos\alpha \geqslant 0.$$

Hence $\sigma$ is decreasing on $\left[0, \frac{\alpha}{2}\right]$, and $\sigma(x) \leqslant \sigma(0) = \rho(\alpha)$.

LEMMA 3. *If $\alpha_1, \ldots, \alpha_n$ are nonnegative and $\Sigma_1^n \alpha_i = \frac{\pi}{2}$, then $\Sigma_1^n \rho(\alpha_i) \leqslant \rho\left(\frac{\pi}{2}\right)$.*

We prove inductively that if $\alpha_i \geqslant 0$ and $\Sigma_1^m \alpha_i \leqslant \pi/2$, then $\Sigma_1^m \rho(\alpha_i) \leqslant \rho(\Sigma_1^m \alpha_i)$. It is clear for $m = 1$, and the step from $m - 1$ to $m$ is a direct application of Lemma 2.

*Proof of Favard's first inequality for symmetric convex sets.* It is an easy consequence of Lemma 3, for we have

$$4\sum_{i}^{n}\sin\frac{\theta_i}{2}-\sum_{i}^{n}\sin\theta_i=2\sum_{1}^{n}\rho\left(\frac{\theta_i}{2}\right)\leqslant 2\rho\left(\frac{\pi}{2}\right)=4.$$

**6. The greatest constant $C$: first method, by Favard's first inequality**

THEOREM 1 [5]. *For every symmetric compact convex set, we have*

$$\Sigma\geqslant\frac{\pi}{(\pi-2)^2}\left(\frac{L}{2}-D\right)^2; \tag{9}$$

*for the euclidean ball, we have equality.*

This means that $\pi/(\pi-2)^2$ is the greatest constant which satisfies inequalities (5) and (8). Note that $\pi/(\pi-2)^2\sim 2.4106\ldots>C_1$.

We shall use a critical but elementary lemma:

LEMMA 4. *Define the function $u$ on $[0,\pi]$ by $\sin x=x-x^3u(x)$. Then $u$ is (strictly) decreasing on $[0,\pi]$.*

Its proof is obvious if we remark that, by Taylor's integral formula,

$$u(x)=\frac{1}{2}\int_0^1(1-t)^2\cos tx\,dt.$$

Other elementary proofs of this result exist, of course.

*Proof of Theorem* 1. We have to prove it for a symmetric inscribed polygon only. Let $S_0=\pi D^2/8$ and $l_0=\pi D/8$ (case of the circle), and let

$$\Delta=\frac{S}{(l-D)^2}-\frac{S_0}{(l_0-D)^2}.$$

Then

$$\Delta=\frac{1}{(l-D)^2(l_0-D)^2}\left\{S\left[(l_0-D)^2-(l-D)^2\right]-(l-D)^2(S_0-S)\right\};$$

but, because $l_0\geqslant l\geqslant D$, we have

$$\Delta\geqslant\frac{1}{(l-D)^2(l_0-D)^2}\left[2S(l_0-l)(l-D)-(l-D)^2(S_0-S)\right].$$

By Favard's first inequality, $2S\geqslant D(l-D)$; hence

$$\Delta\geqslant\frac{1}{(l_0-D)^2}\left[D(l_0-l)-(S_0-S)\right].$$

Inserting the expression for $l$ and $S$ in terms of the angles $\theta_i\geqslant 0$, $\Sigma_1^n\theta_i=\pi$, we have

$$\Delta\geqslant\frac{D^2}{8(l_0-D)^2}\left[4\sum_1^n\left(\theta_i-2\sin\frac{\theta_i}{2}\right)-\sum_1^n(\theta_i-\sin\theta_i)\right]$$
$$=\frac{D^2}{8(l_0-D)^2}\sum_{i=1}^n\theta_i^3\left[u\left(\frac{\theta_i}{2}\right)-u(\theta_i)\right].$$

By Lemma 4, $u\left(\frac{\theta_i}{2}\right)\geqslant u(\theta_i)$; hence $\Delta\geqslant 0$, and

$$\frac{S}{(l-D)^2}\geqslant\frac{S_0}{(l_0-D)^2};$$

this last number is $\pi/2(\pi-2)^2$, and it follows that

$$S \geqslant \frac{\pi}{(\pi-2)^2}\frac{(l-D)^2}{2}.$$

REMARK 2. For a symmetric polygon, the inequality is strict; this follows from the proof of Proposition 2 and Lemma 4.

**7. The greatest constant $C$: direct proof by variational method.** Variational methods are often used in the proof of inequalities about convex sets (see [2] for instance). We shall give a variational proof of Theorem 1, which does not depend on Favard's first inequality. This is a natural proof, but it uses calculus, because the point where an extremum is attained is not completely determined.

We shall prove that if $0 < C < \pi/(\pi-2)^2$, $\theta_i \geqslant 0$, $\Sigma_1^n\theta_i = \pi$, then the quantity

$$\Psi_n(\theta_1,\theta_2,\ldots,\theta_n) = \sum_1^n \sin\theta_i - 4C\left(\sum_1^n \sin\frac{\theta_i}{2} - 1\right)^2 = \frac{8}{D^2}\left[S - \frac{C}{2}(l-D)^2\right]$$

is not negative. This is obvious for $n = 1$, and for $n = 2$ it is equivalent to the inequality

$$\frac{\pi}{(\pi-2)^2} \leqslant \operatorname*{Inf}_{0\leqslant\theta\leqslant\pi} \frac{2\sin\theta}{4\left[\sin\frac{\theta}{2}+\cos\frac{\theta}{2}-1\right]^2} = \frac{1}{2} + \operatorname*{Inf}_{0\leqslant\theta\leqslant\pi}\frac{1}{\sqrt{2}\sin\left(\frac{\theta}{2}+\frac{\pi}{4}\right)-1},$$

i.e., $\dfrac{\pi}{(\pi-2)^2} \leqslant \dfrac{1}{6-4\sqrt{2}}$, and this is true. Hence, *we suppose $n \geqslant 3$.*

Let

$$K_n = \left\{(\theta_1,\ldots,\theta_n)\,|\,\theta_i \geqslant 0,\ \ \sum_1^n\theta_i = \pi\right\},$$

and

$$\mathring{K}_n = \{(\theta_1,\ldots,\theta_n) \in K_n\,|\,\theta_i > 0 \quad \forall i\};$$

it is clear that $\operatorname{Inf}_{\mathring{K}_n}\Psi_n \leqslant 0$ $\quad(\theta_1 = \pi, \theta_i = 0 \quad\text{for}\quad i \geqslant 2)$. We shall prove the following lemma:

LEMMA 5. *If $\Psi_n$ attains its greatest lower bound on $\mathring{K}_n$, for $n \geqslant 3$, this lowerbound is strictly positive.*

*Proof of Theorem* 1 *from Lemma* 5. As $\operatorname{Inf}_{\mathring{K}_n}\Psi_n \leqslant 0, \Psi_n$ will attain its minimum on $K_n$ at a point of $\partial K_n = K_n \setminus \mathring{K}_n$. Thus this minimum will be the minimum of $\Psi_{n-1}$ on $K_{n-1}$. By continuing this procedure, we come to the minimum of $\Psi_2$ on $K_2$, which is positive, and $\Psi_n \geqslant 0$, for all $n$.

*Proof of Lemma* 5. If $\Psi_n$ attains its greatest lower bound on $\mathring{K}_n$ at $(\theta_1,\ldots,\theta_n)$, we have

$$d\Psi_n = \lambda\sum_1^n d\theta_i,\ \text{i.e.,}\ \cos\theta_i - 4C\cos\frac{\theta_i}{2}\left(\sum_1^n\sin\frac{\theta_i}{2} - 1\right) = \lambda,$$

or

$$2\cos^2\frac{\theta_i}{2} - 4C\mu\cos\frac{\theta_i}{2} - \lambda - 1 = 0 \quad \left(\text{where } \mu = \sum_1^n \sin\frac{\theta_i}{2} - 1\right).$$

It follows that there are only two possible values for the $\theta_i$, $\alpha$ and $\beta$: $K$ times $\alpha$ and $(n-K)$ times $\beta$. We may suppose that we have $0 \leqslant \alpha \leqslant \pi/n \leqslant \beta \leqslant \pi$. In such a point, we have

$$\Psi_n = K\sin\alpha + (n-K)\sin\beta - 4C\left[K\sin\frac{\alpha}{2} + (n-K)\sin\frac{\beta}{2} - 1\right]^2.$$

REMARK 3. If we knew that $\alpha = \beta = \pi/n$, then we should have $\Psi_n > 0$ at the point $(\pi/n, \dots, \pi/n)$ for every $C < \pi/(\pi-2)^2$ if and only if it were true that

$$\frac{n\sin\frac{\pi}{n}}{4\left(n\sin\frac{\pi}{2n} - 1\right)^2} \geqslant \frac{\pi}{(\pi-2)^2} \quad \text{for} \quad n \geqslant 2.$$

This inequality is easy to prove with Lemma 4.

In any case, by putting $t = K/n$, $0 \leqslant t \leqslant 1$, Lemma 5 will be true if we show that $A_n \geqslant \pi/(\pi-2)^2$, where

$$A_n = \frac{n[t\sin\alpha + (1-t)\sin\beta]}{4\left[n\left(t\sin\frac{\alpha}{2} + (1-t)\sin\frac{\beta}{2}\right) - 1\right]^2}.$$

With the function $u$ of Lemma 4, we have

$$A_n = \frac{\pi - n[t\alpha^3 u(\alpha) + (1-t)\beta^3 u(\beta)]}{4\left[\frac{\pi}{2} - 1 - \frac{n}{8}\left(t\alpha^3 u\left(\frac{\alpha}{2}\right) + (1-t)\beta^3 u\left(\frac{\beta}{2}\right)\right)\right]^2}.$$

But we have

$$K\sin\frac{\alpha}{2} + (n-K)\sin\frac{\beta}{2} - 1 = \frac{1}{D}(l - D) \geqslant 0,$$

and, if we let

$$X = n[t\alpha^3 u(\alpha) + (1-t)\beta^3 u(\beta)] = \pi - K\sin\alpha - (n-K)\sin\beta \leqslant \pi,$$

then

$$\frac{\pi}{2} - 1 - \frac{X}{8} \geqslant 0.$$

From Lemma 4, $u\left(\frac{\alpha}{2}\right) \geqslant u(\alpha)$ and $u\left(\frac{\beta}{2}\right) \geqslant u(\beta)$.

Hence

$$A_n \geqslant \frac{\pi - X}{4\left[\frac{\pi}{2} - 1 - \frac{X}{8}\right]^2},$$

and it is sufficient to prove that this quantity is greater than $\pi/(\pi-2)^2$, or

$$(*) \qquad X \leqslant \frac{8}{\pi}(\pi-2)(4-\pi).$$

The proof of Lemma 5 will be complete after studying two cases.

1. *First case.* $n - K \geqslant 2$.

As $\beta \leqslant \frac{\pi}{n-K}$ and $|u(x)| \leqslant \frac{1}{6}$, it follows that

$$X \leqslant n\frac{\pi^3}{n^3}\cdot\frac{1}{6} + (n-K)\frac{\pi^3}{(n-K)^3}\cdot\frac{1}{6},$$

and it suffices to see that

$$\frac{1}{n^2} + \frac{1}{4} \leqslant \frac{48}{\pi^4}(\pi - 2)(4 - \pi),$$

or

$$\frac{1}{n^2} \leqslant \frac{48}{\pi^4}(\pi - 2)(4 - \pi) - \frac{1}{4} = \gamma$$

or $n \geqslant \dfrac{1}{\sqrt{\gamma}} \sim 2.072\ldots$, and this is true for $n \geqslant 3$.

2. *Second case.* $n - K = 1$.

We have then $\pi = (n-1)\alpha + \beta$. Three possibilities exist for $\beta$:

(a) $\underline{\dfrac{\pi}{2} \leqslant \beta \leqslant \dfrac{8\pi}{9}}$.

We have

$$X = \pi - (n-1)\sin\alpha - \sin\beta = \pi - \frac{8S}{D^2}.$$

The relation $(*)$ to check is hence

$$\frac{8S}{D^2} \geqslant \pi - \frac{8}{\pi}(\pi - 2)(4 - \pi) = \delta = 0.646\ldots.$$

But $8S/D^2 \geqslant 2\sin\beta$, and it suffices to have $\pi - \beta \geqslant \arcsin\dfrac{\delta}{2}$, or

$$\frac{\pi}{2} \leqslant \beta \leqslant \pi - \arcsin\frac{\delta}{2} = 2.81\ldots,$$

and this is true, for $\dfrac{8\pi}{9} = 2.79\ldots$.

(b) $\underline{\beta \leqslant \dfrac{\pi}{2}}$.

We wish still to have $8S/D^2 \geqslant \delta$; but $8S/D^2 \geqslant (n-1)\sin\alpha$, and $(n-1)\alpha \geqslant \pi/2$; hence

$$\frac{8S}{D^2} \geqslant (n-1)\frac{2}{\pi}\frac{\pi}{2(n-1)} = 1,$$

which is greater than $\delta$ (note that $\alpha \leqslant \pi/2$).

(c) $\underline{\beta \geqslant \dfrac{8\pi}{9}}$.

We have

$$\Psi_n(\alpha, \alpha, \ldots, \alpha, \beta) \geqslant \sin\beta - 4C\left(\sin\frac{\beta}{2} + \frac{\pi - \beta}{2} - 1\right)^2,$$

which will be strictly positive if

$$\frac{\sin h}{4\left[\cos\frac{h}{2} + \frac{h}{2} - 1\right]^2} \geqslant \frac{\pi}{(\pi - 2)^2},$$

where $\beta = \pi - h$, $0 \leqslant h \leqslant \dfrac{\pi}{9}$. But

$$0 \leqslant \cos\frac{h}{2} + \frac{h}{2} - 1 \leqslant \frac{h}{2},$$

hence the first number is greater than

$$\frac{\sin\frac{\pi}{9}}{\frac{\pi}{9}}h \cdot \frac{1}{4\left(\frac{h}{2}\right)^2} = 9\sin\frac{\pi}{9}\cdot\frac{\pi}{h},$$

which will be greater than $\pi/(\pi-2)^2$ if

$$h \leqslant \frac{(\pi-2)^2}{\pi^2} 9\sin\frac{\pi}{9},$$

or

$$\beta \geqslant \pi - \frac{(\pi-2)^2}{2} 9\sin\frac{\pi}{9} = 2.735\ldots,$$

and this is true if $\beta \geqslant 8\pi/9 = 2.79\ldots$.

As pointed out in Remark 3, this proof would be much shorter if the following question had a positive solution:

PROBLEM 1. Among the symmetric convex polygons inscribed in the circle, with $2n$ sides, is the quantity $\dfrac{\Sigma}{\left(\frac{L}{2}-D\right)^2}$ minimum for the regular polygon?

We will use Theorem 1 to give a partial solution to a problem of Favard about the maximum of $L$ when $D$ and $\Sigma$ are given.

**8. Favard's problem on maximality of the perimeter of a convex set.** In [**2**], Favard sets the problem: if $D$ and $\Sigma$ are given, what is the maximum possible for $L$, and for which convex set is it attained? There is no exact general solution, but Favard proved that the maximum, if $\Sigma < \pi D^2/4$, is attained for a polygon inscribed in the circle of diameter $D$, which has all its sides equal, except for one. But the calculation of $L$ may be difficult.

As an example, he proves Favard's first inequality, which states that $L \leq 2D + 2\frac{\Sigma}{D}$. But this inequality is not sharp, for if $\Sigma = \pi D^2/4$, it gives only $L \leqslant \left(2+\frac{\pi}{2}\right)D$, which is without interest, for $L \leqslant \pi D$ in every case.

With Theorem 1, we have another inequality:

$$L \leqslant 2D + 2\frac{\pi-2}{\sqrt{\pi}}\sqrt{\Sigma}. \tag{10}$$

This inequality is sharp for large values of $\Sigma$ for the bound is at most $\pi D$, with equality for $\Sigma = \pi D^2/4$. But, in the case where Favard's first inequality is useful, i.e., if $\Sigma \leq \frac{\pi-2}{2}D^2$, then it is better in the range $0 \leq \Sigma \leq \frac{(\pi-2)^2}{\pi}D^2$.

Hence, we can assert an improved bound for the symmetric case of Favard's problem.

PROPOSITION 3. *For a compact convex symmetric set, we have*

$$2D \leqslant L \leqslant 2D + 2\min\left(\frac{\Sigma}{D}, \frac{\pi-2}{\sqrt{\pi}}\sqrt{\Sigma}\right). \tag{11}$$

**9. Favard's maximal triangle lemma and second inequality.** Favard's maximal triangle lemma is the following:

THEOREM 2 [**2**]. *Consider the convex curves $\Gamma$ contained in triangle $ABC$, with $\hat{C} \geqslant \hat{B}$. Among*

*those that have endpoints B and C and that delimit with BC a given area, the one that has the greatest length is a triangle $BA_0C$, where $A_0$ lies on CA* (see Fig. 6).

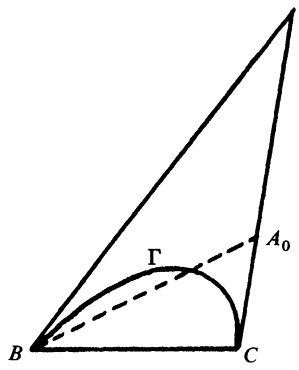



FIG. 6

Favard uses his lemma, for instance, to prove that if a convex set in the circle of diameter $D$ has the greatest perimeter for a given area, then the arcs that bound it inside of this circle are necessarily polygonal, because they are locally polygonal: between two nearby support lines, the boundary is a triangle (make a figure!).

Favard's proof was of a geometric nature. We give proof with functional ideas.

We shall first determine, by functional methods, the extreme points of a convex set of concave functions.

Let $C_0([0, a])$ be the space of continuous functions on $[0, a]$ null at 0 and $a$.

PROPOSITION 4. *Let K be the set of concave continuous functions f on $[0, a]$ satisfying $0 \leqslant f(x) \leqslant \mathrm{Min}(px, q(a-x))$, where $p, q > 0$. Let $S \in ]0, S_0[$, where $S_0 = \dfrac{a^2}{2}\dfrac{pq}{p+q}$, and consider the hyperplane $H_S$ of $C_0([0, a])$ defined by $H_S = \{f \mid \int_0^a f\,dt = S\}$. Then the set $K_S = K \cap H_S$ is a compact convex set in the topology of uniform convergence and its extreme points are functions of the form*

$$f(x) = \mathrm{Min}[px, q(a-x), ux + v, u'x + v'],$$

*for $v \geqslant 0$, $v' \geqslant 0$, $ua + v \geqslant 0$, $u'a + v' \geqslant 0$.*

Recall that a point $f_0$ is extreme in $K_S$ if it is not possible to have $f_0 = (f_1 + f_2)/2$ unless $f_0 = f_1 = f_2 (f_1, f_2 \in K_S)$. We have then eight possible forms for extreme points of $K_S$: see Fig. 7.

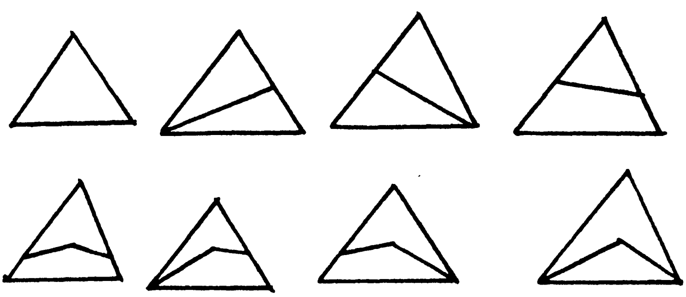


FIG. 7

*Proof.* (a) The set $K_s$ is norm closed and bounded and the functions $f$ in $K_S$ are $\mathrm{Max}(p, q)$-lipschitzian, hence equicontinuous. By Ascoli's theorem, $K_S$ is a compact convex set.

(b) It is almost obvious that the functions $\mathrm{Min}[px, q(a-x), ux+v, u'x+v']$, with the prescribed conditions, are extreme in $K_S$, and we do not give this easy proof.

(c) Define the mapping $T$ from $K_S$ into the set of positive Radon measures on $[0,a]$ by $Tf = -f''$, where $f''$ is the second derivative of $f$ in distribution sense, or equivalently, the Stieltjes' measure $df'_d$ associated to the nonincreasing function $f'_d$, the right derivative of $f$. The measures $\mu$ that are in the range of $T$ must satisfy three conditions, easy to establish:

1. $\int_0^a (a-s)\,d\mu(s) \leqslant p$ (i.e., $f'(0) \leqslant p$ and $\int_0^a f'(t)\,dt = 0$);
2. $\int_0^a s\,d\mu(s) \leqslant q$ (i.e., $f'(a) \geqslant -q$);
3. $\int_0^a \frac{s(a-s)}{2}\,d\mu(s) = S$ (i.e., $\int_0^a f(t)\,dt = S$).

Hence, $\|\mu\| \leqslant p+q$. Let $L$ be the convex compact set of positive measures on $[0,a]$ satisfying $\|\mu\| \leqslant p+q$, with $\mu(\{0\}) = \mu(\{a\}) = 0$, for the topology of duality with the space $C_0([0,a])$, and let be $E_p$, $E_q$ and $M_S$ the closed half-spaces and hyperplane defined by the conditions 1, 2, and 3. Then, the range of $T$ is the set

$$L_S = L \cap E_p \cap E_q \cap M_S,$$

and $T$ is affine and one-to-one from $K_S$ on $L_S$. Thus, the set $\mathscr{E}(K_S)$ of extreme points of $K_S$ is exactly the set $T^{-1}[\mathscr{E}(L_S)]$, where $\mathscr{E}(L_S)$ is the set of extreme points of $L_S$. But we know the set of extreme points of $L$: they are exactly the measures $\|p+q\|\delta_x$, for $x \in ]0,a[$, and the measure 0.

Suppose that we have proved the following lemma:

LEMMA 6. *Let $L$ be a compact convex set in a locally convex Hausdorff topological vector space $X$, and $E_1,\dots,E_k$ closed half-spaces of $X$, $M_1,\dots,M_l$ closed hyperplanes of $X$. Then every extreme point of the convex set $\tilde{L} = L \cap E_1 \cap \cdots \cap E_k \cap M_1 \cap \cdots \cap M_l$ is the barycenter of at most $k+l+1$ extreme points of $L$.*

Then every extreme point $f$ of $K$ is the barycenter of at most four functions $f_1,\dots,f_4$ such that $f''_i = -\lambda_i \delta x_i$, with $\lambda_i \geq 0$; thus, the graph of $f$ is a convex polygon with at most four vertices. But it is easy to see that if two consecutive vertices are inside the triangle made with the lines $y = px$, $y = q(a-x)$, $y = 0$, then the function $f$ is not extreme in $K_S$ (see Fig. 8). Thus, the only possibilities are those described in the proposition.

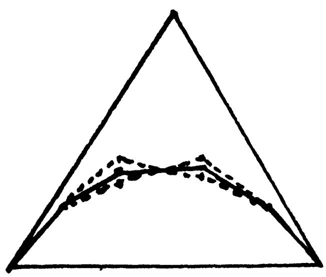


FIG. 8

*Proof of Lemma* 6. Let $x \in \mathscr{E}(\tilde{L})$; if $x$ is not a barycenter of at most $k+l+1$ extreme points of $L$, then $x$ is inside a $(k+l+1)$-dimensional simplex $S$ included in $L$. But $E_1 \cap \cdots \cap E_k \cap M_1 \cap \cdots \cap M_l$ then contains a subspace of dimension at least one, whose intersection with $L$ contains a segment $]x_0, x_1[$ that includes the point $x$. Thus, $x$ cannot be extreme in $\tilde{L}$.

*Proof of Favard's maximal triangle lemma from Proposition* 4. If the angle $\hat{C}$ is greater than $\pi/2$, we change the triangle $ABC$ into $ABC'$, with $\hat{C}' < \pi/2$ and $AB > AC$, adding to the area $S$ and the length $l$ numbers $S_0$ and $l_0$; the problem is not changed and it suffices to solve it in the new situation (see Fig. 9).

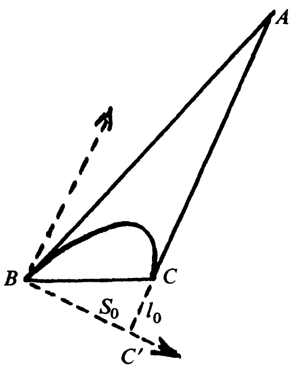



FIG. 9

In the new triangle, convex curves are graphs of functions of $K_S$. Let $l: f \rightsquigarrow l(f)$ be the functional defined by the length of the graph. Then $l$ is a convex continuous function by Lemma 1. We use now a well-known result:

LEMMA 7 (Bauer's maximum principle). *Any convex upper semicontinuous function $l$ on a compact convex set $X$ in a locally convex Hausdorff linear space attains its maximum on an extreme point of $X$.*

The shortest proof comes from [4]: $\{x \in X | l(x) < \mathrm{Max}_X l\}$ is a convex relatively open set, hence its complement in $X$ contains an extreme point of $X$.

Applying Proposition 4, we see that $l$ is maximum on a function of $K_S$ with polygonal graph, which has 1, 2 or 3 vertices. It is not possible that there is a vertex inside the triangle. Otherwise, if $M$ is such a vertex, the others being $F_1$ and $F_2$, the parallel line to $F_1F_2$ through $M$ intersects $BA$ and $CA$ in $C_1$ and $C_2$; one of these points, say $C_2$, is outside the ellipse with foci $F_1$ and $F_2$ passing through $M$; hence $C_2F_1 + C_2F_2 > MF_1 + MF_2$, but the areas of the triangles $F_1MF_2$ and $F_1C_2F_2$ are the same, a contradiction to the maximality of $l$ (see Fig. 10).

Thus, the function in $K_S$ whose graph has a maximum length is polygonal, with two vertices on $BA$ and $CA$, or is a triangle with its vertex on $BA$ or $CA$. But an obvious geometrical argument, which is in [2], shows the triangle $BA_0C$ has maximal length, for if $M_1A_0$ is parallel to $BM_2$ and $\hat{B} \le \hat{C}$, then $\widehat{M_1BM_2} \le \widehat{A_0M_2B}$, and $BM_1 + M_1M_2 \le BA_0 + A_0M_2$ (see Fig. 11).

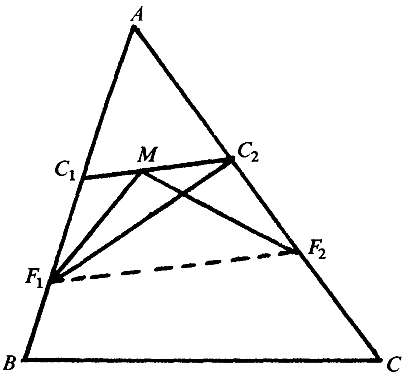



FIG. 10

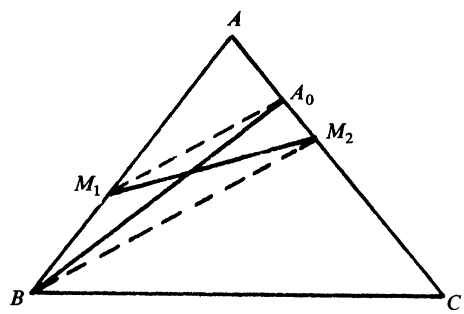



FIG. 11

*Proof of Favard's second inequality.* By density, it suffices to work with concave function $f$ satisfying

$$f(x) \leqslant \mathrm{Min}[px, q(D - x)],$$

and then let $p$ and $q$ tend to infinity. If we take $S$ fixed, $l$ is less than the maximum, given by

Favard's maximal triangle lemma.

But an elementary calculation gives $h = 2S/D$ (see Fig. 12); hence we have

$$l \leqslant h\left[\sqrt{1+\frac{1}{q^2}}+\sqrt{1+\left(\frac{D}{h}-\frac{1}{q}\right)^2}\right],$$

or

$$lD \leqslant 2S\left[\sqrt{1+\frac{1}{q^2}}+\sqrt{1+\left(\frac{D^2}{2S}-\frac{1}{q}\right)^2}\right].$$

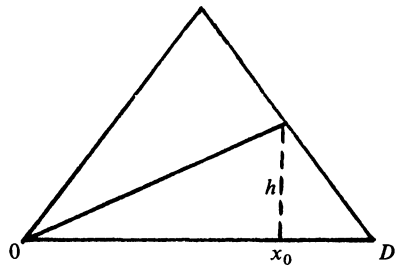



FIG. 12

If $q$ tends to infinity, we obtain the result:

$$lD \leqslant 2S + \sqrt{4S^2 + D^4}.$$

**10. Conclusion and problems.** We recall the first problem already set:

PROBLEM 1. Is the quantity $\dfrac{\Sigma}{\left(\frac{L}{2}-D\right)^2}$ minimum for the regular symmetric polygons, among symmetric inscribed polygons with $2n$ sides?

In connection with Remark 2 we also pose the following problem:

PROBLEM 2. If, for a symmetric compact convex set, we have the equality

$$\Sigma = \frac{\pi}{(\pi-2)^2}\left(\frac{L}{2}-D\right)^2,$$

is this set the euclidean ball?

Many other fascinating problems exist about plane convex sets. The methods for studying them are various: geometrical or analytic, elementary or functional analytic. We hope that the reader will be encouraged to look at some of the problems that are set in the many papers cited in [**1**], [**2**], or [**3**].

### References


1. T. Bonnesen and W. Fenchell, Theorie der konvexen Körper, Chelsea, New York, 1948.
2. J. Favard, Problèmes d'extrémums relatifs aux courbes convexes, Ann. Ecole Norm., 46 (1929) 345–369.
3. H. Hadwiger, Vorlesungen über Inhalt, Oberfläche und Isoperimetrie, Springer, 1957.
4. C. Leger, Convexes compacts et leurs points extrémaux, Acad. Sc., Paris, 267, Série A, 1968, pp. 92–93.
5. M. Rogalski and J. Saint Raymond, Une inégalité concernant les convexes de $\mathbb{R}^2$, Séminaire d'initiation à l'Analyse, 1983–84. Publications de l'Université Pierre et Marie Curie, Paris.

# MODIFIED ITERATION AND PROBABILITY

LAWRENCE J. WALLEN
*Department of Mathematics, University of Hawaii, Honolulu, HI 96822*

It is always a pleasant event when apparently disparate areas of mathematics intersect: when a theorem in one discipline translates into, proves, generalizes, simplifies or "explains" a theorem in another. This phenomenon occurs frequently between probability theory and various aspects of classical analysis. Examples are so abundant that it is perilous even to begin a list. This note presents another such interaction.

Iteration of functions produces fixed points, at least under suitable conditions. We are going to present a modification of the usual iteration procedure which has the pernicious property of retarding or even destroying the convergence to the fixed point and ascertain (purely analytically) just when this convergence is destroyed. Our modification comes about not purely by malice, but turns out to correspond to picking a non-optimal strategy in a certain iterative game of chance. The analytic criterion then reappears as an immediate corollary to an elementary pointwise convergence theorem.

Here is the analytic set-up. $G(w)$ is a strictly convex, continuously differentiable function on the interval $0 \leqslant w \leqslant 1$, with $G(0) > 0$, $G(1) = 1$, and $0 \leqslant G'(w) < 1$ for $0 \leqslant w < 1$. These conditions are more than enough to ensure that if $\eta_{n+1} = G(\eta_n)$ $(\eta_0 = 0)$, then $\eta_n \to 1$, the fixed point of $G$.

Let $w_n$ be a sequence with $0 \leqslant w_n \leqslant 1$. Modify the $\eta_n$ to get a sequence $\xi_n$ by defining $\xi_0 = 0$ and

$$\xi_{n+1} = G(w_{n+1}) + (\xi_n - w_{n+1})G'(w_{n+1}), \tag{†}$$

that is, $(\xi_n, \xi_{n+1})$ lies on the tangent line to $G$ through $(w_{n+1}, G(w_{n+1}))$. The strict convexity of $G$ implies that $\xi_{n+1} \leqslant G(\xi_n)$ with equality exactly when $w_{n+1} = \xi_n$. Since $G$ is a strictly increasing, $\xi_n \leqslant \eta_n$. The scheme reduces to ordinary iteration when $w_{n+1} \equiv \xi_n \equiv \eta_n$.

The question arises, for which sequences $\{w_n\}$ does $\xi_n \to 1$? It is clear pictorially that $w_n$ must tend to 1. In the case where $G'(1) = 1$ (the only interesting one), if we pick $w_{n+1} = 1$, then $\xi_{n+1} = \xi_n$ and no progress has been made. We expect therefore that $\xi_n \to 1$ only when $w_n$ doesn't tend to one too rapidly. This is exactly what happens as the following theorem shows.

THEOREM 1. $\xi_n \to 1$ *if and only if* $w_n \to 1$ *and*

$$\sum_{n=0}^{\infty} (1 - G'(w_n)) = \infty.$$

*Proof.* Assume $1 - \xi_n \to 0$. We want to show the series diverges or, equivalently, $\prod_1^\infty G'(w_n) = 0$. Now

$$\begin{aligned} 1 - \xi_{n+1} &= 1 - G(w_{n+1}) - (\xi_n - 1 + 1 - w_{n+1})G'(w_{n+1}) \\ &= 1 - [G(w_{n+1}) + (1 - w_{n+1})G'(w_{n+1})] + (1 - \xi_n)G'(w_{n+1}). \end{aligned}$$

Owing to the convexity of $G$, the term in brackets is majorized by 1 so that

$$(1 - \xi_{n+1}) \geqslant (1 - \xi_n)G'(w_{n+1})$$

---

*Lawrence J. Wallen*: My vital statistics read as follows: B.A. from Lehigh University, Ph.D. from M.I.T., stints at Bell Laboratories, Stevens Institute of Technology, The University of Michigan, and, finally, the University of Hawaii. I was lured into operator theory (in part) by the seductive lectures on ergodic theory given by the present MONTHLY editor. Most of my research has involved operators, often semigroups of them. Of late, my interests have become more diffuse, my present passion involving some connections between functional analysis and geometry. I watch birds, fish, bugs, stars, sometimes play the fiddle in quartets and sometimes play basketball (in quintets).

and

$$(1-\xi_{n+1}) \geqslant (1-\xi_0)\prod_{j=1}^{n+1} G'(w_j),$$

therefore $\prod_1^\infty G'(w_n) = 0$.

For the sufficiency, since $G(w) > w$, we get

$$G(w) - wG'(w) \geqslant w(1-G'(w)).$$

Pick $N$ so large that $w_n > 1-\varepsilon$ for $n \geqslant N$. Then for these $n$,

$$\begin{aligned}
\xi_n - \xi_{n-1}G'(w_n) = G(w_n) - w_nG'(w_n) &\geqslant (1-\varepsilon)(1-G'(w_n))\\
\xi_{n-1} - \xi_{n-2}G'(w_{n-1}) &\geqslant (1-\varepsilon)(1-G'(w_{n-1}))\\
\vdots \qquad\qquad &\qquad\qquad \vdots\\
\xi_N - \xi_{N-1}G'(w_N) &\geqslant (1-\varepsilon)(1-G'(w_N)).
\end{aligned}$$

Now multiply the second inequality by $G'(w_n)$, the third by $G'(w_n)G'(w_{n-1}),\dots$ and the last by $G'(w_n)G'(w_{n-1})\cdots G'(w_{N+1})$. Adding, we see that both sides telescope and we get

$$\xi_n - \xi_{N-1}\prod_{j=N}^{n} G'(w_j) \geqslant (1-\varepsilon)\left[1-\prod_{j=N}^{n} G'(w_j)\right].$$

Letting $n \to \infty$, we get $\liminf \xi_n \geqslant 1-\varepsilon$ since $\prod_N^\infty G'(w_j) = 0$.

The theorem yields an instant corollary.

COROLLARY 1. (a) $\Sigma(1-G'(\eta_n)) = \infty$ *and*
(b) $\Sigma(1-\eta_n) = \infty$ *if* $G''(w) \leqslant M$ *for* $0 \leqslant w < 1$ *and* $G'(1) = 1$.

*Proof*. For (a), simply set $w_{n+1} = \eta_n$ and use the fact that $\xi_n = \eta_n \to 1$. For (b), the mean value theorem yields $1 - G'(\eta_n) \leqslant M(1-\eta_n)$ and then use (a).

We now make the promised probabilistic connection. The iterative game proceeds as follows. We announce at the outset a sequence of numbers $w_1, w_2, \dots, 0 \leqslant w_n \leqslant 1$. Pick a number $X_1$ randomly from the unit interval, $X_1$ having the continuous distribution $F(x)$. We win $Y_1$ which is $X_1$ if $X_1 > w_1$ and 0 if $X_1 \leqslant w_1$. Now pick $X_2$ (using the same $F(x)$) and win $Y_2$ where $Y_2 = X_2$ if $X_2 > w_2$ and $Y_2 = Y_1$ otherwise. Continue the process ad infinitum.

More rigorously, $\{X_n\}_{n=1}^\infty$ are independent, identically distributed random variables with continuous distribution $F(x)$ satisfying $F(0) = 0$, $F(1) = 1$ and $F$ is strictly increasing on $[0,1]$. $\{Y_n\}_{n=0}^\infty$ is a sequence of random variables defined by $Y_0 = 0$,

$$Y_{n+1} = \begin{cases} X_{n+1} & \text{if } X_{n+1} > w_{n+1},\\ Y_n & \text{if } X_{n+1} \leqslant w_{n+1}.\end{cases}$$

The $w_n$'s are called thresholds. Note they are picked before the game starts.

An ancient question concerns the way to pick $w_1, w_2, \dots, w_n$ so as to maximize $E(Y_n)$ ($E$ is expectation). Compute $E(Y_{n+1})$ as follows:

$$E(Y_{n+1}|X_{n+1} = \xi) = \begin{cases} \xi & \text{if } \xi > w_{n+1},\\ E(Y_n) & \text{if } \xi \leqslant w_{n+1}.\end{cases}$$

Hence

$$E(Y_{n+1}) = \int_0^{w_{n+1}} E(Y_{n+1}|X_{n+1} = \xi)\,dF(\xi) + \int_{w_{n+1}}^1 E(Y_{n+1}|X_{n+1} = \xi)\,dF(\xi)$$

$$= E(Y_n)F(w_{n+1}) + \int_{w_{n+1}}^{1} \xi\, dF(\xi).$$

Set $E(Y_n) = \xi_n$ and $G(w) = 1 - \int_w^1 F(\xi)\,d\xi$. Integrating the last equation by parts gives

$$\xi_{n+1} = G(w_{n+1}) + (\xi_n - w_{n+1})G'(w_{n+1})$$

since $G' = F$. Since $F$ is strictly increasing, we see that $G$ is exactly the sort of function we considered at the outset. The maximum value of $E(Y_n)$ is, since $\xi_n \leqslant \eta_n$, precisely $\eta_n$ and is achieved by picking the $w_{j+1} = \eta_j = E(Y_j)$, $j = 0, \ldots, n-1$. Thus if we play infinitely often, our winnings $(Y_n)$ tend to one.

If, however, we don't play optimally for some reason (there may be additional strictures on choosing the $w_n$'s), we can still ask when our winnings tend to 1.

The following theorem answers the question completely and affords the promised probabilistic proof of Theorem 1.

THEOREM 2. *Still assuming $w_n \to 1$, the following are equivalent:*

(a) $Y_n \to 1$ *with probability one (a.s.).*
(b) $E(Y_n) \to 1$.
(c) $\Sigma(1 - G'(w_j)) = \Sigma(1 - F(w_j)) = \infty$.

*Proof.* Let $(\Omega, P)$ be the probability space underlying the process. For $\omega \in \Omega$, let

$$N'(\omega) = \{\nu\colon X_\nu(\omega) > w_\nu\}.$$

The Borel-Cantelli lemma [1, p. 41] shows that $N'$ is almost surely infinite or finite according as

$$\sum P(X_n > w_n) = \sum(1 - F(w_n))$$

diverges or converges.

For each $n$, let $\nu(n)$ be the largest member of $N'$ satisfying $\nu(n) \leqslant n$ (set $X_0 = 0$, $\nu(n) \equiv 0$ if $N'(\omega) = \varnothing$). Then $Y_n = X_{\nu(n)}$. Thus, when

$$\sum\big(1 - F(w_j)\big) = \infty,$$

we get convergence to one a.s. since $w_n \to 1$ whereas when

$$\sum\big(1 - F(w_j)\big) < \infty,$$

$\lim Y_n(\omega) = X_M(\omega)$ where $M = \max\{\nu\colon \nu \in N'(\omega)\}$ and so the bounded convergence theorem yields

$$\lim E(Y_n) = E(X_M) < 1.$$

Note that in any case $\xi_n = E(Y_n)$ converges.

Actually, this proves Theorem 1 only in the case where $G'(0) = 0$ and $G'(1) = 1$ but the other cases are easily disposed of anyway.

Suppose now we are allowed to choose the $w_n$ (and the associated $Y_n$) using hindsight. That is, take

$$w_n = w_n(X_1, X_2, \ldots, X_{n-1})$$

and define $Y_n$ as before. Then, there is an obvious best strategy. Namely, define

$$\hat{w}_n = \max(X_1, \ldots, X_{n-1})$$

and let $\hat{Y}_n$ be the corresponding $Y_n$, (starting with $\hat{Y}_0 = 0$, $\hat{w}_1 = 0$). A trivial induction argument shows that

$$\hat{Y}_n = \max(X_1, \ldots, X_n).$$

If $\{w_n, Y_n\}$ is any other choice, then since each $Y_n(\omega)$ is one of the $X_j(\omega)$, $j = 1, \dots, n$, we have $\hat{Y}_n \geqslant Y_n$ and consequently $E(\hat{Y}_n) \geqslant E(Y_n)$. In particular, $E(\hat{Y}_n) \geqslant \eta_n$.

Now $E(\hat{Y}_n)$ is easily computed (see $(*)$ below) and is given by

$$E(\hat{Y}_n) = 1 - \int_0^1 F^n(w)\, dw$$

so that

$$1 - \eta_n \geqslant \int_0^1 [G'(w)]^n\, dw.$$

This can also be seen analytically by noting that if $G_j$ denotes the $j$th iterate of $G$, then

$$1 - \eta_n = \int_0^1 \frac{d}{dw} G_n(w)\, dw = \int_0^1 \prod_{j=0}^{n-1} G'(G_j(w))\, dw \geqslant \int_0^1 [G'(w)]^n\, dw$$

since $G(w) \geqslant w$. This in turn gives a sharpening of Corollary 1(b), namely,

$$\sum 1 - \eta_n = \infty \text{ provided } \int_0^1 \frac{dw}{1 - G'(w)} = \infty.$$

We conclude by asking how the best a priori choices of the $w_n$, namely $w_n = \eta_{n-1}$, and the corresponding expected values, $\eta_n$, compare with the best strategy using hindsight, namely $E(\hat{Y}_n)$.

For fixed $n$, let $X_n^1, X_n^2, \dots, X_n^n$ be the order statistics associated with the $X$'s. That is,

$$X_n^1 = \min(X_1, \dots, X_n), \quad \text{and} \quad X_n^j = \text{the } j\text{th smallest } X$$

(see **3**, Chapter 9, for a treatment of order statistics). The expectation of $X_n^k$ is given by the formula,

$$(*) \qquad E(X_n^k) = n\binom{n-1}{k-1} \int_0^1 F^{-1}(u)\, u^{k-1}(1-u)^{n-k}\, du.$$

Consider the family of distributions $F_p(w)$ given by

$$F_p(w) = 1 - (1 - w)^p, \quad p > 0.$$

The asymptotic nature of $\eta_n$ for the corresponding $G_p(w)$ is known:

$$1 - \eta_n \sim \left( \frac{1 + \frac{1}{p}}{n} \right)^{1/p}.$$

A neat proof of this is given in (**4**, p. 223). Putting $F_p$ into $(*)$, routine calculations give

$$(**) \qquad E(X_n^k) = n\binom{n-1}{k-1} \int_0^1 \left(1 - (1-u)^{1/p}\right) u^{k-1}(1-u)^{n-k}\, du$$

$$= 1 - \frac{\Gamma(n+1)}{\Gamma\left(n + \frac{1}{p} + 1\right)} \frac{\Gamma\left(n - k + \frac{1}{p} + 1\right)}{\Gamma(n - k + 1)}.$$

Thinking of $n - k$ as fixed, say $n - k + 1 = \theta$, and using the fact that

$$\Gamma(n + z) \sim n^z \Gamma(n) \quad [\mathbf{2}, \text{p. } 212],$$

we look for $k$ so that $1 - E(X_n^k) \sim 1 - \eta_n$ or, using $(*)$ and $(**)$, we try to solve

$$\frac{\Gamma\left(\theta + \frac{1}{p}\right)}{\Gamma(\theta)} = \left(\frac{p+1}{p}\right)^{1/p}.$$

Here are a few solutions (the last three being approximate):

$$p = 1 \text{ (uniform)}, \qquad k = n - 1,$$
$$p = 1/2, \qquad k = n - 1.50,$$
$$p = 1/3, \qquad k = n - 2.08,$$
$$p = 2, \qquad k = n - 0.73.$$

For the uniform distribution, using the best a priori choices, we do surprisingly well: namely as well as the expectation of the second largest $X$; for $p = 1/3$ we do almost as well as the third largest $X$, and so on. The moral seems to be that the utility of hindsight becomes more pronounced as the probability of getting a large number increases.

**References**

1. L. Breiman, Probability, Addison-Wesley, Reading, MA, 1968.
2. E. T. Copson, An Introduction to the Theory of Functions of a Complex Variable, Clarendon Press, Oxford, 1935.
3. S. Karlin, A First Course in Stochastic Processes, Academic Press, New York, 1966.
4. G. Szekeres, Regular iteration of real and complex functions, Acta Math., 100 (1958) 203–258.

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.
*For instructions about submitting Notes for publication in this department see the inside front cover.*

## ANOTHER NOTE ON THE INCLUSION $L^p(\mu) \subset L^q(\mu)$

ALFONSO VILLANI
*Seminario Matematico, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy*

Throughout this note $(\Omega, \mathscr{A}, \mu)$ will be a positive measure space and, for each $p \in (0, \infty]$, $L^p(\mu)$ will denote the space of all $\mathscr{A}$-measurable real functions $f$ on $\Omega$ such that $\|f\|_p < \infty$, where

$$\|f\|_p = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p} \text{ for } p \in (0, \infty) \quad \text{and} \quad \|f\|_\infty = \operatorname*{ess\,sup}_\Omega |f|,$$

and as usual we identify two functions which differ only on a set of measure zero. When endowed with the metric $d_p$ of convergence in $p$th mean, i.e.,

$$d_p(f, g) = \|f - g\|_p \text{ for } p \in [1, \infty] \quad \text{and} \quad d_p(f, g) = \|f - g\|_p^p \text{ for } p \in (0, 1),$$

$L^p(\mu)$ becomes a complete metric space. We obtain a new characterization of the spaces $(\Omega, \mathscr{A}, \mu)$ for which the inclusion $L^p(\mu) \subset L^q(\mu)$ holds. This result simplifies both the conditions and the proofs already given in [**1**] and [**4**].

We begin with a well-known lemma.

LEMMA 1. *Let $p, q \in [1, \infty]$. The set theoretic inclusion $L^p(\mu) \subset L^q(\mu)$ implies that the inclusion map $i: L^p(\mu) \to L^q(\mu)$ is continuous.*

*Proof.* If $f_n \to f$ in $L^p(\mu)$, then $\{f_n\}$ has a subsequence which converges pointwise almost everywhere to $f$; see [**3**], Theorem 3.12. The desired result now follows easily from the Closed

Graph Theorem; see [2], Theorem 2.15. ■

REMARK. Lemma 1 also holds for $p, q \in (0, \infty]$, with the same proof, even though $L^p(\mu)$ is not a normed space for $0 < p < 1$.

Let $\mathscr{A}_0$ denote the collection of all sets $A \in \mathscr{A}$ with positive measure. Then we have

THEOREM 1. *The following conditions on the measure space* $(\Omega, \mathscr{A}, \mu)$ *are equivalent*:

(1) $L^p(\mu) \subset L^q(\mu)$ *for some* $p, q \in (0, \infty]$ *with* $p < q$,

(2) $\inf_{E \in \mathscr{A}_0} \mu(E) > 0$,

(3) $L^p(\mu) \subset L^q(\mu)$ *for all* $p, q \in (0, \infty]$ *with* $p < q$.

*Proof*. (1) $\Rightarrow$ (2). Since $L^p(\mu) \subset L^q(\mu)$ implies $L^{pt}(\mu) \subset L^{qt}(\mu)$ for every $t \in (0, \infty)$, we can assume $p \geqslant 1$. Then $L^p(\mu)$ and $L^q(\mu)$ are normed spaces, and by Lemma 1 there exists a positive constant $k$ such that $\|f\|_q \leqslant k\|f\|_p$ for every $f \in L^p(\mu)$. In particular we have

$$\{\mu(E)\}^{1/q} \leqslant k\{\mu(E)\}^{1/p},$$

and hence $\mu(E) \geqslant k^{pq/(p-q)}$ for every $E \in \mathscr{A}$ with $0 < \mu(E) < \infty$. This proves (2).

(2) $\Rightarrow$ (3). Let $f \in L^p(\mu)$ and let $E_n = \{|f| > n\}$, $n = 1, 2, \ldots$. By Chebyshev's inequality, $\mu(E_n) \to 0$ as $n \to \infty$, hence, by condition (2), there is an index $n_0$ such that $\mu(E_n) = 0$ for $n \geqslant n_0$, i.e., $|f| \leqslant n_0$ $\mu$-a.e. Thus $L^p(\mu) \subset L^\infty(\mu)$, and this easily implies that $L^p(\mu) \subset L^q(\mu)$ for every $q \in [p, \infty]$.

(3) $\Rightarrow$ (1). This is trivial. ■

Let $\mathscr{A}_\infty$ denote the collection of all sets $A \in \mathscr{A}$ with finite measure. Then we have

THEOREM 2. *The following conditions on the measure space* $(\Omega, \mathscr{A}, \mu)$ *are equivalent*:

(1) $L^p(\mu) \supset L^q(\mu)$ *for some* $p, q \in (0, \infty)$ *with* $p < q$,

(2) $\sup_{E \in \mathscr{A}_\infty} \mu(E) < \infty$,

(3) $L^p(\mu) \supset L^q(\mu)$ *for all* $p, q \in (0, \infty)$ *with* $p < q$.

*Proof*. (1) $\Rightarrow$ (2). As in Theorem 1, we can assume $p \geqslant 1$, so by Lemma 1 there is a positive constant $k$ such that $\|f\|_p \leqslant k\|f\|_q$ for every $f \in L^q(\mu)$. It follows that

$$\mu(E) \leqslant k^{pq/(q-p)} \qquad \text{for every } E \in \mathscr{A}_\infty,$$

and hence condition (2) holds.

(2) $\Rightarrow$ (3). Let $f \in L^q(\mu)$ and let

$$E_n = \{1/(n+1) \leqslant |f| < 1/n\}, \qquad n = 1, 2, \ldots.$$

Then

$$\mu(E_n) \leqslant (n+1)^q \int_\Omega |f|^q \, d\mu < \infty \qquad \text{for every } n = 1, 2, \ldots,$$

and hence, by condition (2), $\Sigma_{n=1}^\infty \mu(E_n) < \infty$, because the $E_n$'s are pairwise disjoint. Now for $p < q$ we have

$$\int_\Omega |f|^p \, d\mu = \int_{\{|f| \geqslant 1\}} |f|^p \, d\mu + \sum_{n=1}^\infty \int_{E_n} |f|^p \, d\mu \leqslant \int_\Omega |f|^q \, d\mu + \sum_{n=1}^\infty \frac{1}{n^p} \mu(E_n) < \infty.$$

(3) $\Rightarrow$ (1). This is trivial. ■

**References**

1. J. L. Romero, When is $L^p(\mu)$ contained in $L^q(\mu)$?, this MONTHLY, 90 (1983) 203–206.
2. W. Rudin, Functional Analysis, McGraw-Hill, 1973.
3. ______, Real and Complex Analysis, 2nd ed., McGraw-Hill, 1974.
4. B. Subramanian, On the inclusion $L^p(\mu) \subset L^q(\mu)$, this MONTHLY, 85 (1978) 479–481.

## SPACES WHERE ALL CONTINUITY IS UNIFORM

M. ARALA CHAVES
*Departamento de Matemática, Faculdade de Ciências, 4000 Porto, Portugal*

Elementary topology courses normally include a proof that all continuous functions from a compact metric space to a metric space are uniformly continuous. We abbreviated this by saying that for compact metric spaces all continuity is uniform. The aim of this note is to give several equivalent conditions, which are necessary and sufficient for all continuity to be uniform. It is easy to check that compactness is not such a condition, because necessity fails.

The conditions are stated formally in the following theorem.

THEOREM. *For a metric space $(X, d)$, the following conditions are equivalent:*

1. *Every continuous function from $X$ to any metric space is uniformly continuous;*
2. *every open covering of $X$ has a Lebesgue number;*
3. *for every sequence $(x_n)$ into $X$ which has no convergent subsequence, the only sequences $(x_n')$ in $X$ such that $\lim d(x_n, x_n') = 0$ are those which are almost equal to $(x_n)$, in the sense that $x_n = x_n'$ for all but a finite set of indices;*
4. *for any infinite subset $A$ of $X$ without accumulation points (in $X$), the infimum of the distances between (different) points of $A$ is greater than $0$.*

The following observations will help to explain how the theorem comes about and how the proof is constructed. Conditions 2 and 3 of the theorem were motivated by a careful analysis of two standard proofs of uniform continuity on compact metric spaces. In fact, one of these proofs [**2**; 11.4, p. 234] merely uses the property of compact metric spaces that every open covering has a Lebesgue number, i.e., a number $\delta > 0$ such that each $\delta$-ball is contained in a set of the covering. That property, which is precisely our Condition 2, is strictly weaker than compactness, as can be easily seen by considering an infinite set with the discrete metric. It turns out, in fact, that uniform continuity of all continuous functions is equivalent to the assertion that every open covering has a Lebesgue number. The other proof of the uniform continuity on compact metric spaces [**1**; 3.16.5, p. 58] uses the characteristic property of compact (metric) spaces that every sequence has a convergent subsequence. It is easy to see that the proof still works if we assume the weaker Condition 3. The interesting point is that once more we have a condition—the third one of our theorem—which is not only sufficient, but also necessary for all continuity to be uniform. Condition 4 is a slightly different and perhaps more suggestive version of Condition 3.

*Proof.* We shall prove our theorem by showing that $1 \Rightarrow 4 \Rightarrow 3 \Rightarrow 2 \Rightarrow 1$.

To prove that 1 implies 4, we will begin by assuming the existence of an infinite subset $A$ of $X$ with no accumulation point in $X$ and such that the infimum of the distances between different points of $A$ is zero. The existence of such a set will enable us to define a continuous function from $X$ to $\mathbb{R}$, which is not uniformly continuous. The general lines for the definition of such a function are as follows:

We first construct a locally finite sequence of balls $(B(x_n, R_n))$ such that: (i) $x_n \in A$, for all $n$; (ii) each ball $B(x_n, R_n)$ has at least one point $x_n'$ of $A$ distinct from $x_n$; (iii) the sequence $(R_n)$ of the radii converges to 0. For each $n$, we then define a real function $f_n$ with support contained in $B(x_n, R_n)$ and such that $f_n(x_n) = 1$ and $f_n(x_n') = 0$. These choices can be made in such a way

that each function vanishes on all points $x'_m$. The sum of these $f_n$ is the continuous function $f$ we wanted to define: the fact that $(d(x_n, x'_n))$ converges to zero, and that $f(x_n) \geqslant 1$ and $f(x'_n) = 0$, ensures that $f$ is not uniformly continuous.

Let us now look into the technical details. Let $r_0 = 1$; assuming $r_i > 0$ to be defined, let $R_{i+1} = \min(r_i, 1/(i+1))$. By hypothesis, there exist in $A$ two distinct points $x_{i+1}$ and $x'_{i+1}$, such that $d(x_{i+1}, x'_{i+1}) < R_{i+1}$; we choose a positive $r_{i+1} < d(x_{i+1}, x'_{i+1})$, such that the $r_{i+1}$-balls centered at $x_{i+1}$ and $x'_{i+1}$ intersect $A$ precisely in their centers. Such an $r_{i+1}$ exists, because $x_{i+1}$ and $x'_{i+1}$ are not accumulation points of $A$. Let us now see why the choices made imply that $x_i, x'_i, x_j, x'_j$ are all different for $i \neq j$. In fact $x_i, x'_i$ (and $x_j, x'_j$) were chosen distinct; by symmetry it is then sufficient to show that, if $i < j$, the points $x_i, x_j$ are different. Now, if $i < j$, we have $R_j \leqslant r_i$ and so the ball $B(x_j, r_i)$ will have at least two points of $A$, $x_j$ and $x'_j$; since $B(x_i, r_i)$ has just one point of $A$, we have $x_i \neq x_j$. Let us prove that the sequence $(B(x_n, R_n))$ is locally finite. We first remark that, for each $a \in X$, there exists an $R > 0$, such that $B(a, R) \setminus \{a\}$ does not intersect $A$. If $a = x_{n_0}$ for some $n_0$, then the balls $B(x_n, R_n)$ and $B(a, R/2)$ are disjoint for $n > \max(n_0, 2/R)$. If no term of the sequence equals $a$, then the same assertion is true if we take $n_0 = 1$. This concludes the proof that $(B(x_n, R_n))$ is locally finite. By defining, for $n \in \mathbb{N}$ and $x \in X$,

$$f_n(x) = \max(0, 1 - d(x_n, x)/r_n),$$

we have a continuous real function $f_n$ on $X$ whose support $\overline{B(x_n, r_n)}$ is included in $B(x_n, R_n)$. As the sequence of these supports is locally finite, it makes sense to define $f = \Sigma f_n$ and this function $f$ is continuous. To finish the proof that 1 implies 4, we prove that $f$ is not uniformly continuous. As $d(x_n, x'_n) < R_n \leqslant 1/n$, it is sufficient to check that, for each $n$, $d(f(x_n), f(x'_n)) \geqslant 1$. From $f_n(x_n) = 1$ it results trivially that $f(x_n) \geqslant 1$ (and it is easy to see that in fact $f(x_n) = 1$, although this is not essential to the proof). Let us now show that $f(x'_n) = 0$. For each $i \in \mathbb{N}$, $x_i$ is the only point of $A$ in the ball $B(x_i, r_i)$. Therefore $x'_n$ does not belong to $B(x_i, r_i)$, since it is a point of $A$ distinct from $x_i$ as we have seen. But, as $f_i$ is zero outside $B(x_i, r_i)$, we conclude that $f_i(x'_n) = 0$, so $f(x'_n) = 0$.

To prove that (4) implies (3), let $(x_n)$ and $(x'_n)$ be sequences such that: (i) $(x_n)$ has no convergent subsequence; (ii) $(d(x_n, x'_n)) \to 0$. We want to show that the two sequences are almost equal. We first remark that $(x'_n)$ has no convergent subsequence either. Let $A$ be the union of the ranges of the two sequences. $A$ is an infinite subset of $X$ with no accumulation points (in $X$). Condition (4) applied to this subset $A$ ensures the existence of an $\varepsilon > 0$, such that the distances between different points of $A$ is greater or equal to $\varepsilon$. For this $\varepsilon$, there exists a $k \in \mathbb{N}$, such that $d(x_n, x'_n) < \varepsilon$ if $n > k$. But $d(x_n, x'_n) < \varepsilon$ implies $x_n = x'_n$; so the two sequences are almost equal.

Assuming now that $X$ satisfies Condition (3), let us prove that every open covering of $X$ has a Lebesgue number (Condition (2)). Suppose the contrary: let $(U_i)_{i \in I}$ be an open covering of $X$ with no Lebesgue number. Then, for each $n \in \mathbb{N}$, $1/n$ is not a Lebesgue number of this covering; therefore there exists a ball of radius $1/n - B(x_n, 1/n)$—not included in any of the open sets $U_i$. Let $i_n$ be such that $x_n \in U_{i_n}$; the relation $B(x_n, 1/n) \not\subset U_{i_n}$ implies the existence in $B(x_n, 1/n)$ of an element $x'_n$ distinct from $x_n$. The sequences $(x_n)$ and $(x'_n)$ are not almost equal and $(d(x_n, x'_n))$ converges to zero. To conclude the proof that (3) implies (2) it is sufficient to show that $(x_n)$ has no convergent subsequence. If $a$ were the limit of a subsequence $(x_{n_k})$, we could choose an open set $U_j$ such that $a \in U_j$ and a $\delta > 0$ such that $B(a, \delta) \subset U_j$; there would be a $p \in \mathbb{N}$ such that $n_p > 2/\delta$ and $x_{n_p} \in B(a, \delta/2)$. Therefore we would have $B(x_{n_p}, 1/n_p) \subset B(a, \delta) \subset U_j$, contradicting the way the $x_n$ were chosen.

For the sake of completeness, we reproduce here the usual proof that the second condition implies the first one. Given a continuous function $f$: $X \to Y$ and an $\varepsilon > 0$, let $\delta > 0$ be a Lebesgue number of the open covering $(f^{-1}(B(y, \varepsilon/2)))_{y \in Y}$ of $X$. Each $\delta$-ball in $X$ is included in an open set of this covering, so that the distance between the images of two points of that ball is less than $\varepsilon$; so $d(x, y) < \delta$ will imply $d(f(x), f(y)) < \varepsilon$.

### References

1. J. Dieudonné, Foundations of Modern Analysis, Academic Press, New York, 1960.
2. J. Dugundji, Topology, Allyn & Bacon, Boston, 1970.

**Added in proof:** See the paper "Is every continuous function uniformly continuous?" by F. Snipes, recently published in Math. Magazine (Vol. 57, 1984, 169–173), for bibliographical references on this and related subjects. Some of these papers state conditions similar to our Conditions 3 and 4.

## UNIQUE RIGHT INVERSES ARE TWO-SIDED

WALTER RUDIN*
*Department of Mathematics, University of Wisconsin, Madison, WI 53706*

A theorem may be hard to discover, even though, once discovered, it is easy to prove. The point of this note is to emphasize that completely nonrigorous (some may say nonsensical) reasoning is perfectly acceptable in the discovery stage, and that it may furnish clues that enable one to make a good guess. Proofs can come later.

Let $R$ be an associative ring, not necessarily commutative, with unit element 1. Recall that an element $a$ of $R$ is said to be *invertible* if there exists $b$ in $R$ so that $ab = 1$ and $ba = 1$. The uniqueness of the inverse $b$ is obvious; in fact, if $ab = 1$ and $ca = 1$, then

$$b = (ca)b = c(ab) = c.$$

Consider the following two questions:

(I) If $1 - xy$ is invertible, must $1 - yx$ be invertible?

(II) If the answer to (I) is yes, is there a simple universal relation between these two inverses, i.e., one that holds in every ring?

Let us try to tackle this by thinking of, say, the complex numbers in place of our ring, where a simple relation exists between inverses and geometric series, namely

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

(at least when $|x| < 1$). It may be nonsense to talk of infinite series whose terms are members of our perfectly arbitrary ring $R$, but never mind. Pretend that the inverse $z$ of $1 - xy$ is given by the geometric series

$$\begin{aligned} z &= 1 + (xy) + (xy)^2 + (xy^3) + \cdots \\ &= 1 + xy + xyxy + xyxyxy + \cdots . \end{aligned}$$

If we accept this, then the inverse $w$ of $1 - yx$, if there is one, ought to be

$$w = 1 + yx + yxyx + yxyxyx + \cdots .$$

Having gone this far, we might as well do a bit of factoring (i.e., assume that the distributive laws extend to our possibly nonexistent infinite series), write

$$w = 1 + y(1 + xy + xyxy + \cdots)x,$$

and observe that the series in parentheses is the postulated expansion of $z$. We are thus led to

$$w = 1 + yzx.$$

So far, we have proved *nothing*. But we have found a candidate for $w$, and can test whether it does the job. Indeed,

*Partially supported by the National Science Foundation and by the William F. Vilas Trust Estate.

$$\begin{aligned}(1-yx)w &= (1-yx)(1+yzx)\\ &= 1-yx+yzx-yxyzx\\ &= 1-yx+y[(1-xy)z]x,\end{aligned}$$

which is 1, because the quantity in brackets is 1.

A closer look at this computation shows something else: we have only used the assumption that $z$ is a *right* inverse of $1-xy$ and have deduced that $w$ is then a *right* inverse of $1-yx$. The same is true with left in place of right, because

$$(1+yzx)(1-yx) = 1-yx+y[z(1-xy)]x.$$

Thus we get the following result, which actually does more than just answer our two questions:

THEOREM 1. *If $z$ is a right [left] inverse of $1-xy$, then $1+yzx$ is a right [left] inverse of $1-yx$.*

As we just saw, the proof of this theorem is a total triviality. But if we had been unwilling to use infinite series in a context where they may make no sense, how difficult would it have been to *discover* $1+yzx$?

When $y=1$, our two questions are of course pointless, but Theorem 1 tells us something even then that we might not have noticed otherwise, namely:

*If $z$ is a right inverse of $1-x$, so is $1+zx$.*

Let us now assume that $1-x$ has a unique right inverse $z$. Then it follows that $1+zx=z$. This implies $z(1-x)=1$, *so that $z$ is also a left inverse of $1-x$!* In other words, $1-x$ is invertible. Since every element of $R$ can be written in the form $1-x$, we have arrived at the result to which the title of this note alludes.

THEOREM 2. *The invertible elements of $R$ are precisely those that have unique right inverses in $R$.*

Of course, the same is true with left in place of right.

The following well-known fact from linear algebra is also an immediate consequence of Theorem 1:

*If $A$ and $B$ are $n$-by-$n$ matrices over some field, then $AB$ and $BA$ have the same eigenvalues.*

P.S. After completing this note, I was told that the geometric series trick of finding $1+yzx$ is described in [**1**] and [**2**]. However, no conclusions about one-sided inverses are drawn there. In the context of Banach algebras (where series do make sense) $1+yzx$ occurs in an exercise on p. 259 of [**4**]. The referee has pointed out that Theorem 2 appears as Exercise 6 on p. 89 of [**3**].

**References**

1. P. R. Halmos, Does mathematics have elements? The Math. Intelligencer, 3 (1981) 147–153.
2. ______, Does mathematics have elements? Bull. Austral. Math. Soc., 25 (1982) 161–175.
3. N. Jacobson, Basic Algebra, vol. 1, Freeman, 1974.
4. W. Rudin, Functional Analysis, McGraw-Hill, 1973.

---

**ANSWER TO PHOTO ON PAGE 465**

David Blackwell.

CENTER SECTION

(Vol. 92, No. 7, August-September 1985)

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, P.** AAAS Report X: Research & Development, FY 1986. AAAS, 1985, xviii + 270 pp, (P). [ISBN: 0-87168-272-9] Tenth annual analysis of the federal R&D budget request prepared by AAAS. Includes chapters on agency budgets (NSF, DOD, Energy, etc.) and discipline analyses. In this budget, mathematics was recommended for a 15% increase at NSF, tilted towards young investigators. LAS

**General, S, P*, L**.** Metamagical Themas: Questing for the Essence of Mind and Pattern. Douglas R. Hofstadter. Basic Books, 1985, xxviii + 852 pp, $24.95. [ISBN: 0-465-04540-5] A magnificent celebration of pattern and metaphor with imaginative and thought-provoking examples drawn from science, language, art, and international affairs. Contains 25 Hofstadter columns from Scientific American plus eight other chapters tied together with extensive new post-thema notes. Includes a lengthy annotated bibliography of incredible variety, and a thorough index. LAS

**General, S*, L*.** Atoms of Silence: An Exploration of Cosmic Evolution. Hubert Reeves. MIT Pr, 1984, xii + 244 pp, $14.95. [ISBN: 0-262-18112-6] An extraordinary, poetic portrait of the history of the universe, of the "evolution" of quarks to atoms to molecules to cells to primates to man. Astrophysicist Reeves uses metaphor to trace nuclear evolution, chemical evolution, and, briefly, biological evolution in a masterful synthesis of scientific knowledge. LAS

**General, S.** Beat the Racetrack. William T. Ziemba, Donald B. Hausch. Harcourt Brace Jovanovich, 1984, xx + 392 pp, $22.95. [ISBN: 0-15-111275-4] Using the assumption (based on strong empirical evidence) that the probability that a horse will win a race can be closely estimated by the relative amount bet on the horse, the authors indicate how to take advantage of underbets that a horse will place or show (finish in the top two or top three respectively) to get a positive expected gain. RSK

**Education, S(15-16).** Word Problems for Maxima and Minima from Computations to Equations. Stanley Bezuszka. Motivated Math Project Activity, Booklet 13. Boston College Pr, 1984, 240 pp, (P). All the max-min problems a teacher would ever want to see. Problems solved without calculus by writing computer programs to list all instances and looking for the maximum or minimum values. Emphasis on developing the algebraic equation through computation and observation of patterns. Useful as source of exercises. MW

**History, S*, P**, L**.** I Want To Be A Mathematician: An Automathography. Paul R. Halmos. Springer-Verlag, 1985, xvi + 421 pp, $41.50. [ISBN: 0-387-96078-3] 400 pages of unadulterated Halmos on teaching ("The hardest part...is to keep your mouth shut"), on scholarship ("the most important attribute of a genuine professional mathematician"), on writing ("I like words more than numbers...I want to say that as an indication of mathematical ability, liking words is better than being good at calculus"), on research ("The greatest kind of step forward is the illuminating central example from which it is easy to get insight into all the surrounding sweeping generalities"), and on himself: "My greatest strength as a mathematician is the ability to see when two things are the ´same´." LAS

**History, T*(15-16: 1), S*, L*.** The History of Mathematics, An Introduction. David M. Burton. Allyn & Bacon, 1985, ix + 678 pp, $40. [ISBN: 0-205-08095-2] Unencyclopedic. For college juniors and seniors. On the level of Eves´ popular history of mathematics. Accessible to general reader. Covers 5000 years through the early part of the 20th century. Some emphasis on personalities. Problems at various levels of difficulty are integral to the text and appear throughout. Author´s "personal tastes and prejudices held sway" in selection of material. Numerous references, many to recent papers and articles, are given in each chapter. Suggestions for further reading. Well-illustrated. Altogether readable. JK

**Logic, S(16-18), P*, L**.** Cantorian Set Theory and Limitation of Size. Michael Hallett. Logic Guides, No. 10. Clarendon Pr, 1984, xxii + 343 pp, $32.50. [ISBN: 0-19-853179-6] A comprehensive study of the impact that Cantor´s metaphysical, philosophical, and theological views had on his sub-

stitution of sets for numbers as the foundation of mathematics. Constructivist and "creational" metaphors in which sets exist in Divine intellect explain Cantor's thesis that "possibility implies existence;" Aristotelian and scholastic teachings on the actual infinite led Cantor to posit an Absolute Infinite beyond the ken of human intellect. Corrects a number of common misunderstandings and establishes a new plateau for historical comprehension of Cantor's monumental contribution to mathematics. LAS

**Combinatorics, P.** Finite Generalized Quadrangles. S.E. Payne, J.A. Thas. Research Notes in Math., No. 110. Pitman, 1984, 312 pp, $22.95 (P). [ISBN: 0-273-08655-3] Combinatorial and geometric study of a type of point-line incidence structure introduced by J. Tits. General structure, combinatorial characterizations, embeddings in affine and projective spaces, collineation groups, coset geometries. Analysis of "small" examples. RM

**Number Theory, T(18: 1), S, P.** Arithmetic Functions and Integer Products. P.D.T.A. Elliott. Grund. der math. Wissenschaften, B. 272. Springer-Verlag, 1985, xv + 461 pp, $64. [ISBN: 0-387-96094-5] The aim of this volume is to introduce a new aesthetic into the study of arithmetic functions by systematically applying them to certain problems in algebra. For example, the problem of representing an integer as a product to certain rationals touches upon the theory of denumerably infinite groups. An impressive volume with many new results. Exercises, unsolved problems, and an extensive bibliography. CEC

**Number Theory, P.** Lecture Notes in Mathematics-1082: Arithmetik Abelscher Varietäten mit komplexer Multiplikation. Claus-Günther Schmidt. Springer-Verlag, 1984, x + 96 pp, $7.50 (P). [ISBN: 0-387-13863-3]

**Number Theory, S(17), P*, L.** Exercises in Number Theory. D.P. Parent. Problem Books in Math. Springer-Verlag, 1984, x + 541 pp, $38. [ISBN: 0-387-96063-5] A translation of the 1978 French edition. A collection of about 200 problems and solutions from 10 different areas of number theory. Each section includes a brief introduction to the area. The combined work of a group of French number theorists. A valuable resource. CEC

**Group Theory, S(18), P.** Lecture Notes in Mathematics-1084: Structure and Representations of Q-Groups. Dennis Kletzing. Springer-Verlag, 1984, vi + 290 pp, $13.50 (P). [ISBN: 0-387-13865-X] Investigation of the structure of rational groups (finite groups whose complex representations have rationally valued characters, e.g., Weyl groups) and the relationship between the rationally represented characters and the permutation characters. Uses techniques based on algebraic geometry (local characters, Burnside rings, etc.). RM

**Topological Groups, P.** Lecture Notes in Mathematics-1079: Compact Semitopological Semigroups: An Intrinsic Theory. Wolfgang Ruppert. Springer-Verlag, 1984, 260 pp, $13.50 (P). [ISBN: 0-387-13387-9] Presents major areas of current research on compact semitopological semigroups and semigroup compactifications. Emphasis is on the topological-algebraic structures involved; methods from functional analysis are avoided as much as possible. DA

**Topological Groups, P.** Transformation Groups on Manifolds. Ted Petrie, John D. Randall. Pure & Appl. Math., V. 82. Dekker, 1984, vii + 266 pp, $49.75. [ISBN: 0-8247-7156-7] Gives a discussion of some aspects in the study of transformation groups of manifolds. It features equivariant K-theory, new material on equivariant surgery, and Smith equivalence. LN

**Algebra, T(18), S, P.** Lecture Notes in Mathematics-1081: Modular Representation Theory: New Trends and Methods. David Benson. Springer-Verlag, 1984, xi + 231 pp, $11 (P). [ISBN: 0-387-13389-5] These lecture notes consist of two chapters. The first gives background material from the theory of rings and modules, and the second treats representation rings, almost split sequences and the Auslander-Reiten quiver, and complexity and cohomology varieties in detail. The material covered has little overlap with existing texts. Includes exercises and references. CEC

**Calculus, T(13: 2).** BASIC Fundamentals and Style. James S. Quasney, John Maniotes. Boyd & Fraser, 1984, xvi + 463 pp, $19 (P). [ISBN: 0-87835-138-8] A BASIC programming text which emphasizes principles of structured programming. Many skill exercises; excellent programming problems. Good chapter summaries and self-test exercises. MA

**Calculus, T*(13: 1).** Calculus: A Modeling Approach, Third Edition. Marvin L. Bittinger. Addison-Wesley, 1984, xi + 559 pp, $31.95. [ISBN: 0-201-11217-5] For a one-term course in calculus. Applications in business, economics, the behavioral sciences, the social sciences, biology and medicine. Changes include availability of computer software keyed to the text, video tape reviews for self-study and a student's guide to the exercises. "Challenge" problems have been added. Approach remains "experience based." Leisurely-paced and readable with nice graphics. If a course you offer fits, have a look. (First Edition, TR, June-July 1976; Second Edition, TR, January 1981.) JK

**Complex Analysis, T(18: 2).** Compact Complex Surfaces. W. Barth, C. Peters, A. Van de Ven. Erg. der Math. und ihrer Grenzgebiete, 3. Folge, Band 4. Springer-Verlag, 1984, x + 304 pp, $38. [ISBN: 0-387-12172-2] A nearly complete theory of compact complex surfaces, including a nice historical introduction. LN

**Complex Analysis, P.** Proceedings of the Conference on Banach Algebras and Several Complex Variables. Ed: F. Greenleaf, D. Gulick. Contemp. Math., V. 32. AMS, 1983, xiii + 295 pp, $28 (P). [ISBN: 0-8218-5034-2] Twenty-six papers constituting the proceedings of a conference on Banach algebras and

several complex variables, held at Yale in June, 1983. The papers treat Banach algebras, function algebras, and finite and infinite dimensional holomorphy. PZ

**Partial Differential Equations, T(18: 1, 2), S.** Maximum Principles in Differential Equations. Murray H. Protter, Hans F. Weinberger. Springer-Verlag, 1984, x + 261 pp, $28. [ISBN: 0-387-96068-6] A comprehensive introduction to the idea of maximum principle as it occurs in partial differential equations. Four chapters develop the theory in the settings of one dimension, elliptic equations, parabolic equations, and hyperbolic equations. With exercises and an extended bibliography. Reprint of 1967 Prentice-Hall edition (TR, April 1968). PZ

**Partial Differential Equations, T*(15-16: 1), S, L.** Boundary-Value Problems. Ladis D. Kovach. Addison-Wesley, 1984, xii + 419 pp, $33.95. [ISBN: 0-201-11728-2] Separation of variables and transform (Laplace and Fourier) methods for solving second-order linear partial differential equations which arise in applications. Light on theory but honest. References provided for omitted proofs. Exercises to clarify the text and to fill in omitted details, to provide practice in using the methods and to extend the theory. Some are of a computational nature. Numerous footnotes with historical tidbits. Readable, non-overpowering, of manageable length in one course and student-friendly. JK

**Functional Analysis, S(18), P.** Boundary Methods: An Algebraic Theory. Ismael Herrera. Applic. Math. Ser. Pitman, 1984, viii + 136 pp, $32.95. [ISBN: 0-273-08635-9] An outgrowth of work on diffraction problems that occur in geophysics and seismic engineering. The algebraic theory is formulated in the setting of general functional-valued operators defined on arbitrary linear spaces which do not generally possess an inner product or a metric. Employs Trefftz procedure and Green's formulas. MU

**Analysis, S(18), P.** On a New Method of Analysis and Its Applications. Paul Turán. Wiley, 1984, xvi + 584 pp, $49.95. [ISBN: 0-471-89255-6] The author's theory of "power sums" (linear combinations of powers of complex numbers) arose from his study of the Riemann hypothesis. In Part I, the minimax theory of power sums is developed. Part II is applications, principally to analytic number theory, but also to numerical analysis, complex analysis, and differential equations. There are two sections on unsolved problems. PZ

**Analysis, T(18: 1, 2), S*, P.** Exercises in Integration. Claude George. Problem Books in Math. Springer-Verlag, 1984, x + 550 pp, $38. [ISBN: 0-387-96060-0] Eleven chapters of graduate-level exercises, each with a clear and detailed solution. Nearly all integrals are with respect to Lebesgue measure on $R^n$. Fourier analysis figures prominently. Zeroth chapter outlines the necessary theory, without proofs. PZ

**Analysis, P.** The Geometric Theory of Ordinary Differential Equations and Algebraic Functions. Georges Valiron. Transl: James Glazebrook. Lie Groups: History, Frontiers & Applic., V. XIV. Math Sci Pr, 1984, xiii + 576 pp, $75. [ISBN: 0-915692-38-4] This classic treatise of analysis provides a fascinating exposition of Riemann surfaces, elliptic integrals, Weierstrass' preparation theorem in two complex variables, as well as a detailed study of differential equations and of the calculus of variations. Bad typing has disfigured this work, however--everywhere numerous lacunae stand as ghosts of missing clauses, absent symbols make formulae meaningless, and even characters have undergone odd permutations. Valiron's masterpiece deserved better care. YN

**Algebraic Geometry, P.** Lecture Notes in Mathematics-1056: Algebraic Geometry, Bucharest 1982. Ed: L. Bădescu, D. Popescu. Springer-Verlag, 1984, 380 pp, $16 (P). [ISBN: 0-387-12930-8] Contains twelve research papers in algebraic geometry which were presented at an August 1982 conference held at the University of Bucharest. LN

**Algebraic Geometry, T(16-18: 2).** Geometry of Projective Algebraic Curves. Makoto Namba. Pure & Appl. Math., V. 88. Dekker, 1984, x + 409 pp, $83.50. [ISBN: 0-8247-7222-9] Discusses algebraic plane curves including a classification of the singularities of curves of low degree. After introductory material on general varieties, the monograph gives most of the classical theory of curves without using sheaf cohomology. LN

**Algebraic Geometry, P.** Lectures on Results on Bezout's Theorem. W. Vogel. Springer-Verlag, 1984, 132 pp, $7.30 (P). [ISBN: 0-387-12679-1] The main result shows that the product of the degrees of two intersecting subvarieties of projective n-space can be expressed in terms of degrees of subvarieties of the intersection and lengths of primary ideals. This generalizes results on Bezout and is valid even if the intersection is excess. LN

**Algebraic Geometry, T(18: 2), P.** Isolated Singular Points on Complete Intersections. E.J.N. Looijenga. London Math. Soc. Lect. Note Ser., No. 77. Cambridge U Pr, 1984, xi + 200 pp, $22.95 (P). [ISBN: 0-521-28674-3] An organized text on isolated singularities of complete intersections. Contains a treatment of Milnor fibrations, monodromy, critical spaces, deformations, and the Gauss-Manin connection. LN

**Algebraic Geometry, P.** Introduction to Intersection Theory in Algebraic Geometry. William Fulton. CBMS Reg. Conf Ser. in Math., No. 54. AMS, 1984, v + 82 pp, $16 (P). [ISBN: 0-8218-0704-8] An introduction to the modern theory of excess intersection theory of algebraic varieties. Gives a readable presentation with many examples and applications. LN

**Algebraic Geometry, P.** Cohomology of Quotients in Symplectic and Algebraic Geometry. Frances Clare Kirwan. Math. Notes, No. 31. Princeton U Pr, 1984, 210 pp, $17.50 (P). [ISBN: 0-691-08370-3] Kirwan develops a general procedure for computing the rational cohomology of quotients of groups acting on nonsingular complex projective varieties. She has two approaches--one algebraic and one based on Morse theory and symplectic geometry. LN

**Differential Geometry, S(18), P.** Automorphic Forms and the Picard Number of an Elliptic Surface. Peter F. Stiller. Aspects of Math., V. E5. Friedr. Vieweg & Sohn (US Distr: Heyden & Son), 1984, 194 pp, $14 (P). [ISBN: 3-528-08587-8] Deals with the question of how many curves there are on an elliptic surface up to algebraic equivalence. Includes a large number of examples for which the Picard number is explicitly determined. Includes new results. References. CEC

**Differential Geometry, P.** Harmonic and Minimal Maps: With Applications in Geometry and Physics. Gâbor Tŏth. Ser. in Math. & Its Applic. Halsted Pr, 1984, v + 342 pp, $70. [ISBN: 0-470-20127-4] Studies the geometry of harmonic and minimal maps into spaces of constant curvature. Gives applications to physics and geometry. LN

**Geometry, S*, P**, L**.** Mathematics and Optimal Form. Stefan Hildebrandt, Anthony Tromba. Scientific American Books (Distr: WH Freeman), 1984, xiii + 215 pp. [ISBN: 0-7167-5009-0] A stunning presentation of the geometry of optimal shapes--minimal paths and surfaces, soap films--together with stories of the mathematicians who discovered them. Profusely illustrated with both drawings and photographs, this first mathematics volume in the Scientific American Library is a dazzling masterpiece of exposition and geometry. LAS

**Geometry, P, L.** Geometry of Spatial Forms. Peter C. Gasson. Ser. in Math. & Its Applic. Halsted Pr, 1983, xii + 561 pp, $64.95. [ISBN: 0-470-20011-1] Intended for shape design architects and engineers, this "encyclopaedic coverage of all the important concepts and formulae of applicable geometry" would be extremely useful to anyone interested in computer-aided graphics. JNC

**Algebraic Topology, P.** Lecture Notes in Mathematics-1051: Algebraic Topology, Aarhus 1982. Ed: I. Madsen, B. Oliver. Springer-Verlag, 1984, x + 665 pp, $25.50 (P). [ISBN: 0-387-12902-2] Proceedings of an August 1982 conference at the Mathematics Institute at Aarhus, Denmark. Includes papers in algebraic K- and L-theory, geometry of manifolds, homotopy theory, and transformation groups. LN

**Operations Research, T**(14), L.** A First Course in Mathematical Modeling. Frank R. Giordano, Maurice D. Weir. Brooks/Cole, 1985, xvii + 382 pp, $32.50. [ISBN: 0-53-403367-9] Provides an early modeling experience assuming only one-variable calculus. Discussion of creative and empirical model construction, model fitting, models requiring optimization techniques, and the modeling of dynamic behavior. Excellent examples. Suggestions for projects and for further reading. Highly readable. JRG

**Optimization, T(14-15).** Linear Programming: An Introduction to Finite Improvement Algorithms. Daniel Solow. Elsevier Sci, 1984, xv + 392 pp, $37. [ISBN: 0-444-00912-4] Why another linear programming book? The author's distinctives: greater emphasis on proofs (including an entire chapter on the nature and methods of mathematical proof); a downplay of the tableau method in favor of an emphasis on conceptualization (who does these by hand anyhow?); an account of most recent computational techniques. An attractive book! AWR

**Statistics, S(14-18), P.** Applied Factorial and Fractional Designs. Robert A. McLean, Virgil L. Anderson. Statistics, V. 55. Dekker, 1984, ix + 373 pp, $65. [ISBN: 0-8247-7154-0] About 100 pages of non-theoretical discussion of factorial and fractional designs followed by a 260-page appendix containing three out-of-print publications on designs from the National Bureau of Standard and tables for partially nonorthogonal designs. FLW

**Statistics, T(13-14: 1).** Basic Statistics for Nurses, Second Edition. Rebecca Grant Knapp. Wiley, 1985, xii + 418 pp, $15.95 (P). [ISBN: 0-471-87563-5] This revision of this non-mathematical introduction contains new material on estimation, the power of a test, and experimental design. It also includes a decision tree to help guide one to the appropriate analysis. (First Edition, TR, December 1978.) FLW

**Statistics, T, S(14-15).** Computer Modeling for Business and Industry. Bruce L. Bowerman, Richard T. O'Connell. Statistics: Textbooks & Mono., V. 59. Dekker, 1984, vii + 219 pp, $23.75. [ISBN: 0-8247-7296-2] Self-contained chapters on basic descriptive statistics and the chi-square test, regression analysis, elementary time-series, and linear programming illustrate the use of SAS and LINOD to solve problems in statistics and operations research. Intended as a supplement to business statistics and quantitative analysis courses. JRG

**Statistics, P.** Design of Experiments, Ranking and Selection: Essays in Honor of Robert E. Bechhofer. Ed: Thomas J. Santner, Ajit C. Tamhane. Statistics, V. 56. Dekker, 1984, xxii + 302 pp, $45. [ISBN: 0-8247-7274-1] A biographical essay and bibliography for Robert E. Bechhofer followed by 21 research papers on multiple comparisons, selection and ranking, and estimation and testing. FLW

**Statistics, T*(14-15: 1, 2).** Probability and Statistics for Engineers, Third Edition. Irwin Miller, John E. Freund. Prentice-Hall, 1985, x + 530 pp, $34.95. [ISBN: 0-13-711938-0] Revision of the author's 1977 Second Edition (TR, November 1977). Major changes include an updating of the applications, notation and terminology, a revision of the exercises (particularly in the earlier chapters), new material on combinatorial methods and stem-and-leaf plots, a revised treatment of statistical

inference, and the addition of a check list of key terms at the end of each chapter. RSK

**Statistics, P*.** Lecture Notes in Statistics-25: Time Series Analysis of Irregularly Observed Data. Ed: Emanuel Parzen. Springer-Verlag, 1984, 363 pp, $21 (P). [ISBN: 0-387-96040-6] Proceedings of a symposium held at Texas A&M University in February, 1983. Contains 16 papers by active researchers in this important area. RSK

**Statistics, T(1).** Business Statistics: Concepts and Applications, Second Edition. William J. Stevenson. Harper & Row, 1985, xvii + 637 pp, $31.95. [ISBN: 0-06-046442-9] Probability, sampling, estimation, hypothesis testing, regression, correlation, index numbers and time series. New topics in Second Edition (First Edition, TR, December 1978): multiple regression, decision theory, two-way analysis of variance. Chapters, not sections, are basic units. Informal approach--no proofs. Each chapter includes outline, objectives and review. Study guide available. MA

**Statistics, T(17-18: 1, 2), P*.** Statistical Inference Based on Ranks. Thomas P. Hettmansperger. Wiley, 1984, xvii + 323 pp, $37.50. [ISBN: 0-471-88474-X] In the Wiley Series in Probability and Mathematical Statistics. Major goal is "to develop a coherent and unified set of statistical methods (based on ranks) for carrying out inferences in various experimental situations." Includes one and two-sample location models, one and two-way layouts, a systematic treatment of the general linear model, and a final chapter on multivariate location models. Emphasizes both efficiency and robustness of these methods. Good set of references. RSK

**Statistics, P, L*.** Handbook of Applicable Mathematics. V. VI, Statistics. Ed: Emlyn Lloyd. Wiley, 1984, $85 each [ISBN: 0-471-90024-9]. Part A, xxxviii + 541 pp [ISBN: 0-471-90274-8]; Part B, xxxvi + 486 pp. [ISBN: 0-471-90272-1] Fifth volume to appear of this six-volume series aimed at professional users of mathematics (Volume V: Geometry and Combinatorics has yet to appear). (For description of series, see TR, December 1980, of Volume I: Algebra.) Part A deals with basic principles of estimation and testing. Part B treats linear models, sequential analysis, distribution-free methods, Bayesian statistics, multivariate analysis, time series, decision theory and Kalman filtering. RSK

**Statistics, S(15-17), P.** Applied Statistics: A Handbook of Techniques, Second Edition. Lothar Sachs. Transl: Zenon Reynarowych. Ser. in Statistics. Springer-Verlag, 1984, xxviii + 707 pp, $54. [ISBN: 0-387-90976-1] Modest revision of the author's 1982 First Edition (TR, June-July 1983). Changes include some corrections and insertions in the text, and some additions and deletions from the indices and bibliography. RSK

**Statistics, P*.** The Collected Works of George E.P. Box. Ed: George C. Tiao. Stat. & Prob. Ser. Wadsworth, 1985. Volume I, xiv + 657 pp, $42.95 [ISBN: 0-534-03307-5]; Volume II, xiv + 710 pp, $44.95. [ISBN: 0-534-03308-3] Includes a major portion of his articles. Volume I contains papers in two main areas: statistical inference, robustness, and modeling strategy; and experimental design and response surface methodology. Volume II has three parts: time series analysis and forecasting; distribution theory, transformation of variables, and nonlinear estimation; and applications of statistics. RSK

**Computer Literacy, S, P, L*.** The Sachertorte Algorithm and Other Antidotes to Computer Anxiety. John Shore. Viking, 1985, xvi + 270 pp, $16.95. [ISBN: 0-670-80541-6] A sound, informative and well-written introduction to computers for computer-anxious laymen. Totally avoiding the cute patronizing tone of so many books of this genre, physicist Shore offers a perceptive, wise and balanced view of computers, citing appropriate examples and authorities from both computer science and society. A worthy sequel to Weizenbaum's Computer Power and Human Reason. LAS

**Computer Programming, T(14: 1), S, L.** Problem Solving with Fortran. Richard W. Dillman. Holt, Rinehart & Winston, 1985, xiii + 354 pp, (P). [ISBN: 0-03-063734-1] Introduction to programming using ANSI FORTRAN 77. Use of structured control primitives gives the book much of its strength. Good programming style. Emphasis on stepwise refinement is lacking. FA

**Computer Programming, S(15-16).** MACRO 11 Programming. C.C. Zammit. Adam Hilger (US Distr: Heyden & Son), 1984, xi + 119 pp, $19 (P). [ISBN: 0-85274-769-1] A comprehensive tutorial on Macro-11 assembly language programming. Sections describe the PDP-11 architecture, the assembler, the instruction set, the use of macros, I/O and modular programming. FA

**Computer Programming, T(14-16: 1), S, P, L.** Understanding C. Bruce H. Hunter. Sybex, 1984, xiv + 320 pp, $17.95 (P). [ISBN: 0-89588-123-3] Written for a beginning C programmer (who knows BASIC, at least), this introduction moves quickly through data types and pointers to an extensive treatment of C functions. Intended for the microcomputer (CP/M or MSDOS) world, it includes extensive tables comparing various C compilers. No exercises. LAS

**Software Systems, S, P.** The TK!Solver Book: A Guide to Problem-Solving in Science, Engineering, Business, and Education. Milos Konopasek, Sundaresan Jayaraman. Osborne/McGraw-Hill, 1984, vi + 458 pp, $19.95 (P). [ISBN: 0-88134-115-0] A thorough introduction to "TK!Solver," the most commonly used "tool kit" package for solving algebraic equations. The system stores "rules" relating variables to each other, then solves (directly or by iteration) for any one given adequate data on the others. This package has enormous potential for imaginative pedagogy in high school mathematics. LAS

**Computer Science, P.** Lecture Notes in Computer Science-170: 7th International Conference on Automated Deduction. Ed: R.E. Shostak. Springer-Verlag, 1984, vi + 508 pp, $22.50 (P). [ISBN: 0-387-96022-8] Proceedings of a May 1984 conference held in Napa, California. The keynote address gives an overview of the field. LN

**Computer Science, P.** Lecture Notes in Physics: WOPPLOT 83: Parallel Processing: Logic, Organization, and Technology. Ed: J. Becker, I. Eisele. Springer-Verlag, 1984, v + 189 pp, $10 (P). [ISBN: 0-387-12917-0] Proceedings of a June 1983 workshop held at the Federal Armed Forces University in Munich. LAS

**Computer Science, P.** Synthesis of Digital Designs from Recursion Equations. Steven D. Johnson. MIT Pr, 1984, xv + 207 pp, $30. [ISBN: 0-262-10029-0] Functional languages (Scott-Strachey formalism) and applicative notation are used as basis for digital design. Design process viewed as transformation from specification to realization, with self-referential realization corresponding to feedback in physical implementation. 1983 ACM Distinguished Dissertation. RM

**Computer Science, P.** Automated Theorem Proving: After 25 Years. Ed: W.W. Bledsoe, D.W. Loveland. Contemp. Math., V. 29. AMS, 1984, ix + 360 pp, $30 (P). [ISBN: 0-8218-5027-X] Expanded proceedings of a special session on automated theorem held at the 1983 Denver meeting of AMS. First article is an introduction to and overview of the field. LN

**Computer Science, T(13-14).** Computer Mathematics. D.J. Cooke, H.E. Bez. Computer Sci. Texts, No. 18. Cambridge U Pr, 1984, xii + 394 pp, $49.50; $19.95. [ISBN: 0-521-25341-1; 0-521-27324-2] A textbook to acquaint computer science students with needed mathematical structures and techniques, e.g., algebraic structures, matrices, graph theory, languages and grammars, finite automata, computer geometry. LN

**Applications (Artificial Intelligence), P.** Developments in Expert Systems. Ed: M.J. Coombs. Computers & People Ser. Academic Pr, 1984, xv + 253 pp, $15. [ISBN: 0-12-187580-6] 13 articles on the methods and applications of expert systems, knowledge-based computer systems for solving problems at an expert level, employing casual knowledge, heuristics, user interaction, explanations, set coverings and natural language understanding techniques. Applications include electronic circuits, medical diagnosis, business operations, and processing assistance, biographies and forecasting. RWN

**Applications (Economics), P.** Practical Statistical Sampling for Auditors. Arthur J. Wilburn. Statistics, V. 52. Dekker, 1984, x + 410 pp, $39.75. [ISBN: 0-8247-7124-9] Non-mathematical description of various sampling procedures appropriate for use in auditing, intermixed with discussions of related auditing problems. Includes a presentation of the author's flexible sampling strategy, a Bayesian approach which is particularly suitable in this context. RSK

**Applications (Engineering), T(17).** Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Yosef Sheffi. Prentice-Hall, 1985, xvi + 399 pp, $45.95. [ISBN: 0-13-939729-9] The flow pattern of traffic through an urban network is analyzed by modeling travel decisions and network congestion. The book grew out of notes used in undergraduate and first year graduate courses in traffic management at MIT. AWR

**Applications (Physics), S(18), P.** Mathematical Aspects of Superspace. Ed: H.-J. Seifert, C.J.S. Clarke, A. Rosenblum. NATO ASI Ser. C, V. 132. D Reidel, 1983, xii + 214 pp, $39. [ISBN: 90-277-1805-9] These papers represent an effort to unify and extend the mathematics underlying modern relativity and quantum theory. Such questions are addressed as "Does there exist a superspace having all the properties that physicists require of it?" Very densely written, requiring familiarity with a great deal of mathematics and physics. MU

**Applications (Social Science), T*(15-16: 1), S*, L.** Probability in Social Science. Samuel Goldberg. Math. Modeling, No. 1. Birkhauser Boston, 1983, xii + 119 pp, $18.95. [ISBN: 3-7643-3089-9] Revisions of seven of fifteen CUPM modules originally distributed in an MAA sourcebook in 1977. Topics range from power indices to glottochronology (rate of change of language), and employ primarily undergraduate probability. Each module concludes with careful exercises and an extensive bibliography making contact with current social science research. LAS

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

**Section Reports**

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

## Eastern Pennsylvania and Delaware Section

The annual spring meeting of the Eastern Pennsylvania and Delaware Section was held on March 30, 1985 at Bloomsburg University, Bloomsburg, Pennsylvania. There were approximately 90 registrations.

Invited Lecture:

"Mathematical Education in America," by Garrett Birkhoff, Harvard University.

Workshop:

"Discrete Algorithmic Mathematics," by Stephen B. Maurer, Swarthmore College.

## Florida Section

The eighteenth annual meeting of the Florida Section was held on March 8-9, 1985 at Stetson University in DeLand, Florida. There were 147 registrants.

Invited Addresses:

"Classification of the Simple Groups," by Daniel Gorenstein, Rutgers University.
"Renewing Collegiate Mathematics," by Lynn Arthur Steen, St. Olaf College.
"Some Problems Neither My Computer Nor I Can Solve--YET," by Richard V. Andree, University of Oklahoma.
"What's the Delay? An Introduction to Differential Equations Involving Delays," by Mary Parrott, University of South Florida.
"An Interesting Linear Transformation and Its Generalizations," by Sidney Kung, Jacksonvil University.
"Space-Time Geometry," by Gregory Galloway, University of Miami.
"Aspects of Legislation Affecting Mathematics Education," by T.K. Wetherell, Chair, Committee on Higher Education, Florida House of Representatives.

Short Presentations:

"An Introduction to TEAM, Teaching Experimental Applied Mathematics," by David Lawson, Stetson University.
"Inner Models of Zermelo-Fraenkel Set Theory with Urelements (ZFU) and Standard Models of Arithmetic," by David Lawson, Stetson University.
"A Connected Locally-Connected Countable Space Satisfying the Separation Axiom $T_3 - (1/n)$ for Every n," by Jennifer L. Davidson and Gerhard Ritter, University of Florida.
"Discovering Mathematical Concepts Through the Computer," by Gareth Williams, Stetson University.
"Some Notes on Heron Triangles," by Alan Wayne, Pasco-Hernando Community College.
"Computers for Communication in Mathematics," by Beverly Brechner, University of Florida.
"Matrices That Multiply Like Rabbits," by S.P. Miles, Florida State University.
"Using the Computer in Analysis--Chasing Delta Through Dirichlet's Function on the MacIntosh," by Harold Dansberger, Stetson University.
"Arcade Games, Kidney Stones and Reflections on an Ellipsoid: A Computer Graphics Demonstration in Color," by Arthur Cummer, University of Florida.
"The Geometry and the Group Theory of Special Relativity," by Ernest A. Thieleker, University of South Florida.

Special Sessions:

"Articulation Session," organized by Ernest R. Ross, Jr., St. Petersburg Junior College.
"ICME-5 Panel Discussion," organized by Don Hill, Florida A&M University; Robert Kalin, Florida State University; Don and Betty Lichtenberg, University of South Florida.
"Special Session on Certification," organized by Louis M. Edwards, Valencia Community College; Don Lichtenberg, University of South Florida; Tom Denmark, Florida State University; Clem Boyer, Florida Council of Teachers of Mathematics.
Regional Meeting Session, organized by Elizabeth Magarian, Stetson University.
Florida Two-Year College Mathematics Association Meeting, organized by Dale Grussing, Miami-Dade Community College, North.
Women in Mathematics Meeting, organized by Marilyn Repsher, Jacksonville University: "Education and Examination Policies of the Actuary Societies," by Ruth Frew, American Society of Pension Actuaries.
Florida Association of Mathematical Educators Meeting, organized by Robert Kalin, Florida State University.

Student Papers:

"Two Telescoping Series," by Jan Figa, University of Central Florida.
"Fantastic Feminine Figures," by Kimberly Kailing, University of Tampa.

At the business meeting an award was presented to Dwight B. Goodner of Florida State University for outstanding contributions to mathematics and the mathematical community of the Florida Section.

## Southern California Section

The spring meeting of the Southern California Section was a joint meeting with SIAM, held on March 2, 1985 at California State University Northridge, at Northridge, California. There were 103 registrants.

Invited Addresses:

"Computation of Small Scale Phenomena in the Solution of Differential Equations," by Bjorn Engquist, University of California at Los Angeles.
"Statistical Inference for Discovery Process Models," by Louis Gordon, University of Southern California.
"A Survey of Work on the Bieberbach Conjecture," by Carl Fitzgerald, University of California at San Diego.
"Computational Complexity of Linear Programming," by Kenneth Lane, Colby College.
"The Effect of Battery Additives on Numerical Analysis," by John Todd, Cal Tech; Magnus Hestenes, University of California at Los Angeles.

Panel Discussion:

"Mathematics or Computer Science: What Does Industry Really Want?" by Roy Danchick, TRW (Moderator); Diane Schwartz, CSUN; Stavros Busenberg, Harvey Mudd College; Barney Krinsky, Hughes Aircraft Company.

## Northern California Section

The annual meeting of the Northern California Section was held on February 23, 1985 at Menlo College at Atherton, California. There were 224 registered participants.

Invited Addresses:

"Unsolved Problems in Convexity: Some Very Easy Questions With Very Hard Answers," by David Barnette, University of California at Davis.
"Reminiscences on the Origins of Linear Programming," by George Dantzig, Stanford University.
"Mathematics, Statistics, and Computers," by Bradley Efron, Stanford University.
"Cosets, Spinsters, Clusters, and the Schroeder-Bernstein Theorem," by Paul Halmos, University of Santa Clara.
"A Non-Mathematician's Apology," by Constance Reid.

## Maryland-D.C.-Virginia Section

The fall meeting of the Maryland-D.C.-Virginia Section was held on November 9-10, 1984 at the U.S. Naval Academy in Annapolis, Maryland.

Invited Addresses:

"Shakespeare or Bacon? Submarine or Whale?" by Herbert Wilf, University of Pennsylvania.
"Mathematics and The Art of Escher," by Doris Schattschneider, Moravian College.

Short Presentations:

"Applications of Jacobi's Method for Computing Singular Values," by T. Hoy Booker, Gallaudet College.
"Computers in the Mathematics Classroon: A Good Idea? Maybe," by Ephraim Salins, Montgomery College.
"An Extension of Euler's Theorem," by William Wardlaw, U.S. Naval Academy.
"TEAM (Teaching Experiential Applied Mathematics)," by John Smith, George Mason University.
"Percolation in Continuous Systems," by James R. Kirkwood, Sweet Briar College.
"Teaching a Calculus of Programming," by Harlan D. Mill, IBM and University of Maryland.
"Inertial Flight Instruments: Mathematics Applied to Flight Training Simulations: A Case Study," by Tom Allan, Sperry Corporation.
"$L^2$ on the Unit Circle: Which Rotationally Invariant Subspaces are Hardy-like?" by Richard B. Tucker, Mary Baldwin College.
"Computer Science Curricula: Where's the Math?" by Richard W. Dillman, Western Maryland College.
"Sine, Cosine, and Binomial Coefficients," by Michael Hoffman, U.S. Naval Academy.
"Ring Theory and Exact Integer Arithmetic," by George Mackiw, Loyola College.

"Using the Computer Package SAS in an Undergraduate Course," by Raymond Geremia, Goucher College.
"A Monte Carlo Simulation Involving Alzheimer's Disease," by John C. Hennessey, Loyola College.
"Mathematics Liberally Applied for/by the Non-Mathematics Major," by Sister Helen Christensen, Loyola College.
"Hyperbolic Boundary Value Problems Solved by Special Functions Methods," by Peter McCoy, U.S. Naval Academy.
"Synthesizing the Discrete and Continuous: A Potential First Year College Math Course," by Stephen B. Seidman and Michael D. Rice, George Mason University.
"A Moth to the Flame," by Robert A. Maynard, Tidewater Community College.
"RMS Without Calculus," by Robert A. Maynard, Tidewater Community College.
"An Application Course for the Terminal Mathematics Student," by Thomas Sonnabend, Trinity College.
"Subparticles and Their Mathematical Generation," by Robert A. Herrmann, U.S. Naval Academy.
"Tax Brackets, Loans and Effective Interest Rates," by John Milcetich, University of the District of Columbia.
"Are the Japanese Really Ahead in Mathematics Education?" by Donald R. Peeples, Mary Washington College.
"Pragmatics, Transactions, and Consensic Logics: Applications in Programming Languages and Artificial Intelligence," by John Hays, Naval Research Laboratory.

Student Papers:

"Pursuit Games," by Young Lee, Goucher College.
"Optimum Strategy for a Two Person Poker Game," by Laura Lamb, Goucher College.
"Guessing a Number with Lying," by Susan Emily Imber, Goucher College.

## Louisiana-Mississippi Section

The sixty-second annual joint meeting of the Louisiana-Mississippi Section of MAA and the Louisiana-Mississippi Branch of NCTM was held on February 15-16, 1985 in Biloxi, Mississippi. The University of Southern Mississippi served as host; there were 209 registrants.

Invited Addresses:

"Count or Think," by Paul R. Halmos, University of Santa Clara.
"How to Paint Triangles," by Paul R. Halmos, University of Santa Clara.

Contributed Papers:

"Estimation of the Pareto Density Function," by Badiollah R. Asrabadi, Nicholls State University.
"Hyperbolas of Statistical Confidence," by Margaret W. Maxfield, Louisiana Tech University.
"Comparing Two Estimators of Reliability for the Gamma Failure Model," by Laurie Meaux and Thomas L. Boullion, University of Southwestern Louisiana.
"Comparing Estimators of Reliability for the Uniform Distribution," by Sharon Navard and Thomnas L. Boullion, University of Southwestern Louisiana.
"Unitary Cycles Less Than Four," by Dale Woods, Central State University, Edmond, Oklahoma.
"Equivalent Induced Topologies and Semi-topological Properties," by Charles Dorsett, Louisiana Tech University.
"A Characterization of Infinite Dimension for Vector Spaces," by Henry Heatherly, University of Southwestern Louisiana.
"A General Principle for Fixed Points of Contractive Multi-valued Mappings," by Hideaki Kaneko, Mississippi State University.
"Connected Sums on the Hyperbolic Plane," by Margaret M. LaSalle, University of Southwestern Louisiana.
"Constructing Covering Spaces of 3-manifolds," by William Ortmeyer, University of Southwestern Louisiana.
"Automorphism Groups II," by Gary L. Walls, University of Southern Mississippi.
"A Ring for Quantum Mechanics," by Fred Rolfes, University of Southwestern Louisiana.
"Periods of Sums of Periodic Functions," by James Caveny, University of Southern Mississippi.
"Multisolute Network Flow Problems," by J.B. Garner, Louisiana Tech University.
"A Note on Frobenius Type Solutions of Second Order Linear Differential Equations," by William F. Hoyt, Mississippi State University.
"Sufficient Conditions for Convergence to Zero of Solutions of Second Order Differential Equations," by J.R. Graef and P.W. Spikes, Mississippi State University.
"The House Number Problem and Its Variations," by Joey Paul, Copiah-Lincoln Junior College.
"Some Remarks on Equations of Matrices," by Cecil E. Robinson, Jr., Univ. of Southern Mississippi.
"Remarks on an Open Question in LCP," by M.W. Jeter and W.C. Pye, Univ. of Southern Mississippi.
"Who Takes Science? Characteristics of 342 Students," by Bill Attebery, M. Maxfield, M. Livingston, and T. Portis, Louisiana Tech University and Grambling State University.
"Calculators Are Not Obsolete," by N. Jane Carr, McNeese State University.
* "Consistency," by W. Ed Stephens, McNeese State University.
* "Is It Working? A Five-year Evaluation of the Developmental Mathematics Program at McNeese State University," by Robert T. Yellott, McNeese State University.
"Are Perfect Numbers Really Perfect?" by Steve Ligh, University of Southwestern Louisiana.
"A Necessary and Sufficient Condition for Convergence of a Positive-term Series," by Robert Heller, Mississippi State University.

Student Papers:

"Difference Equations and Probability," by Kate Abernethy, Mississippi University for Women.
"Remarks on the Convergence of $\pi = 4(1-1/3+1/5-1/7+\ldots)$," by Sharon Crook, University of Southern Mississippi.
"The Elastica Problem," by Deanna M. Caveny, University of Southern Mississippi.
"Bezout's Theorem," by Susan M. Gibson, University of Southern Mississippi.
"Disk Storage Efficiency," by Aubrey Griffin, Xavier University of New Orleans.
"The Method of Curvature," by Donna L. Huch, University of Southern Mississippi.
"Power Series Solutions of the Lane-Emden Equation," by Randall A. Stevenson, Northeast Louisiana University.
"Improper Integrals with the Greatest Integer Function," by Carolyn Autrey, University of Southern Mississippi.
"A Number-theoretic Result from Abstract Algebra," by Scott J. Beslin, University of Southwestern Louisiana.
"Covering Theorems for the Alternating Group," by Karen Fawcett, Univ. of Southern Mississippi.
"Some Properties of P-matrices," by Richard Godbold, University of Southern Mississippi.
"A Generalized Thomas-Fermi Theory of Ionized Atoms," by Yiu Chung Hon, University of Southwestern Louisiana.
"Irrational Numbers Viewed as Continued Fractions," by Lida G. McDowell, University of Southern Mississippi.
"An Elementary Solution to Equations Such as $x^2 = x \pmod{p^k}$ and $x^3 = x \pmod{p^k}$, by D. Straub, University of Southwestern Louisiana.

Panel Discussions:

"Everything You've Always Wanted to Know About Challenging Your Bright Students, But Were Afraid to Ask," by Burrow Brooks, Delta State University (Moderator); James Porter, University of Mississippi; Duane Blumberg, University of Southwstern Louisiana; Donald Ryan, Northwestern State University of Louisiana.
"Discrete Mathematics: Something New in the First Two Years of Undergraduate Education?" by Jerome A. Goldstein, Tulane University (Moderator); R.D. Anderson, Louisiana State University; Bettye Anne Case, Florida State University.

Short Course:

"Curves and Surfaces--The Building Blocks in Computer Aided Design and Numerical Simulation," by Wayne Mastin, Mississippi State University.

Special Report:

"Status of the Profession," by Bernard Madison, University of Arkansas.

## Eastern Pennsylvania and Delaware Section

The annual fall meeting of the Eastern Pennsylvania and Delaware Section was held on November 17, 1984 at Swarthmore College. There were 163 persons in attendance.

Invited Addresses:

"Mordell's Conjecture: Ideas at the Confluence of Arithmetic and Geometry," by Stephen S. Shatz, University of Pennsylvania.
"Renewing Undergraduate Mathematics," by Lynn A. Steen, St. Olaf College.
"Solutions of Diophantine Equations and the Class Number Problem of Gauss," by Don Zagier, University of Maryland.
"Remediation: A Waste Or a Gold Mine?" by Raymond Coughlin, Temple University.

## Oklahoma-Arkansas Section

The forty-seventh annual meeting of the Oklahoma-Arkansas Section was held on March 29-30, 1985 at the University of Tulsa at Tulsa, Oklahoma. There were 148 registrants.

Invited Addresses:

"Some Functions That Count," by Gerald L. Alexanderson, University of Santa Clara.
"What is Geometry: Euclid, Riemann, Weil, or Bing?" by Saunders MacLane, University of Chicago.

Short Presentations:

"Need Retraining in Computer Science? I Have an Answer," by Edward N. Mosley, Arkansas College.
"A Study of the Effect of Math and Science Preparation on ACT Scores," by Darryel Reigh, University of Science and Arts of Oklahoma.
"Graphing Absolutely--Taxi and Maxi Geometrics," by Robert DeMille, Holland Hall School.
"Pythagorean Triples and Trigonometric Identities," by Thomas Peter, University of Arkansas at Little Rock.

"Optimizing Gauss-Seidel Iteration," by John M. Woods, Oklahoma Baptist University.
"LOGO for Third and Fourth Graders," by John Watson, Arkansas Tech University.
"Improbable Probabilities with $\pi$ and e," by Robert DeMille, Holland Hall School.
"Sub-surface Electrical Measurements About Plane and Cylindrical Interfaces," by D.J. Boyce and Dale Woods, Central State University.
"A Model for the Bending of a Plastic Trough," by Larry Crawford, Oklahoma State University.
"A Problem of a Perimeter Estimation for an Elliptic Partial Differential Equation," by Kip Holm, University of Oklahoma.
"On the Relationship Between Levinson Recursion and the R and S Arrays for ARMA Model Identification," by Brenda J. Roberts, University of Tulsa.
"Radar Assisted Turns for Helicopter," by James Yates, Central State University.
"Singular Perturbations," by Walter Kelley, University of Oklahoma.
"Lotka-Volterra Competition Equations: Numerically," by Jeff Thompson, Oklahoma State University.
"On Inclusions Between $L^p(\mu)$ Spaces," by Joe Swartz, Oklahoma State University.
"Boundedness of Solutions of Difference Equations and Application to Numerical Solution of Integral Equations," by Zkzislaw Jackiewicx, University of Arkansas.
"Zermelo's First Proof," by Herman Burchard, Oklahoma State University.
"Positive Integral Operation," by Calvin Piston, John Brown University.
"Lotka-Volterra Competition Models Revisited--I," by Marvin S. Keener, Oklahoma State University.
"Lotka-Volterra Competition Models Revisited--II," by Stanley F. Fox and Keith W. Dorschner, Oklahoma State University.
"Population Equilibrium and Stability in Plant-Animal Pollination Systems," by Harrington Wells, University of Tulsa.
"On Groupoids," by Naoki Kimura and Jane Sheu, University of Arkansas.
"A Problem in Linear Algebra Related to Azumaya's Exact Rings," by Joel Haack, Oklahoma State U.
"Lattices of Transformations," by Sinisa Crzenkozic and Michael Dunze, University of Arkansas.
"Topology of p-adic Fields," by Paul T. Young, Oklahoma State University.
"A Taste of Differential Topology," by Bobby N. Winters, Oklahoma State University.
"A Quadratic Transformation on the Plane," by T. Sekiguichi, University of Arkansas.
"A Topological Proof of the Compactness of the Propositional Calculus," by Michelle Penner, Oklahoma State University.
"C[0,1] Has No Minimal Topology As a Riesz Space," by Daryl Ezzo, University of Tulsa.
"Stability Analysis of Linear Multistep Methods for Delay Differential Equation," by V.L. Bakke, University of Arkansas.
"Banach Spaces on Which Every Completely Continuous Operator is Compact," by Tommy Leavelle, John Brown University.
"Sharp Weak-type Inequalities for Analytic Functions on the Unit Disc," by Boguslaw Tomaszewski, Oklahoma State University.
"The Moment Problem and $C^n$-scalar Operators," by Ralph de Laubenfels, University of Tulsa.
"TEAM Materials for the Classroom," by Dan Hansen, Northeastern Oklahoma State University.
"Use of the MAA Placement Exam at ECU," by Ray Hamlett, East Central State University.
"James Clerk Maxwell and the Development of Vector Analysis," by Richard Greenhaw, Oklahoma Christian College.
"How Do Electronic Machines Perform Arithmetic?" by Gary V. Millsap, East Central University.
"Mathematics and Statistics in Medicine and Epidemiology," by Elisa T. Lee, University of Oklahoma Health Sciences Center.
"On Nicholson's Blowflies and Other Bugs," by Carlos Castillo-Chavez, University of Tulsa.
"Stability Results for a Discrete Predator-Prey Model," by William Ray, University of Oklahoma.

Student Papers:

"Strategic Metal Stockpiling," by Steven Cudd and Lisa Pfannestiel, Oklahoma State University.
"Mathematics: Agony or Ecstasy," by Tammy Tucker, Hendrix College.
"Computer Assisted Enrollment," by Donnie J. Glass, University of Science and Arts of Oklahoma.
"A Characterization of a Class of Stone Algebras by Factorization Lattices," by Scott Roberts, Hendrix College.
"Predator-Prey Systems and Optimal Harvesting," by Gordon Clark, Ryan Cross, and David Patocka, Oklahoma State University.
"Applying Prime Factorizations: Amicable and Perfect Numbers," by James Hart, Hendrix College.

## Rocky Mountain Section

The sixty-eighth annual meeting of the Rocky Mountain Section was held March 15-16, 1985 on the campus of Casper College.

Invited Addresses:

"Women Mathematicians: Grace Chisolm Young and Emmy Noether," by Lida Barrett, University of Northern Illinois.
"Bieberbach, Riemann, and the Class of 1990," by Kenneth Gross, University of Wyoming.
"Topology and Material Science," by Lida Barrett, University of Northern Illinois.

Short Presentations:

"A New 'Problem Solving' Course," by Ben Roth, University of Wyoming.
"The Probability of Obtaining All Roots of a Polynomial Equation to be Real (Part 1)," by Hung C.

Li, University of Southern Colorado.

"Getting Beyond the Formulas--Generating Graphical Representations of Fourier Series Solutions for the Heat Equation," by Joan Hundhausen, Colorado School of Mines.

"The Bootstrap with Some Applications," by Martin Hamilton, Montana State University.

"Calculation of Critical Values for the Studentized Range," by Richard Lund.

"Probabilistic Design in Engineering," by Leon Borgman, University of Wyoming.

"Two-sided Distribution-free Tolerance Intervals and Accompanying Sample Size Problems," by William Guenther, University of Wyoming.

"Planning, Implementing and Obtaining Results for a Statewide Mathematics Assessment Test in Montana," by Ken Tiahrt, Montana State University.

"Results on the Structure of the Polynomial Ring k[x]," by Karen Whitehead, South Dakota School of Mines & Technology.

"The Technique of Nonlinear Variation of Constants--Conceptual and Quantitative Results," by Michael Lord, Colorado School of Mines.

"Comments on Two Mathematical Experiences for Non-mathematically Oriented Students, by John Hodges, University of Colorado at Boulder.

"Isomorphisms Can be Non-trivial," by Stephen Bronn, University of Southern Colorado.

"Quarternion Propagation Anomaly," by Thomas Kelly, Colorado School of Mines.

Student Papers:

"Revised Fast Fourier Transforms," by Shane Hubler, Colorado College.

"An Application for Multiple Regression," by Donna Johnson, South Dakota School of Mines & Tech.

"Probabilistic Rocket-engine Propellant Sizing," by Lloyd Best, Colorado School of Mines.

"What Happens to Students Who Start Out with Remedial Math?" by Kim Gehring, South Dakota School of Mines & Technology.

"The Random Walk Program," by Steve Kelly, George Washington High School, Denver, Colorado.

"A New Technique for Identifying Transfer Function-noise Models," by Paul Anderson, Colorado School of Mines.

"Patterns in Bernoulli Numbers," by Shane Hubler, Colorado College.

"Mathematics in Hungary," by Louis Kovari, Colorado School of Mines.

## Kansas Section

The seventieth annual meeting of the Kansas Section was held at Kansas Wesleyan in Saline, Kansas on March 29-30, 1985. There were 80 registrants.

Invited Addresses:

"The Reality of Students' Mathematical Behavior," by Alan H. Schoenfeld, University of California at Berkeley.

"Teaching a Problem Solving Course," by Alan H. Schoenfeld, University of California at Berkeley.

"What High School Seniors Know and How They Learned It," by Curtis McKnight, Univ. of Oklahoma.

Short Presentations:

"On the Existence and Nonexistence of Periodic Orbits in a Small Neighborhood of a Homoclinic Orbit in $R^2$ and $R^3$," by Mohammad Riazal-Kermani, Fort Hays State University.

"More on Super Heronian Triangles," by Ruth Meyer, Kansas State University.

"An Introduction to Strongly Regular Graphs and Association Schemes," by Cecil Andrew Ellard, Kansas State University.

"Practical Teaching Suggestions for Improvement of Problem Solving Skills in Mathematics," by Carolyn Ehr, Fort Hays State University.

"Synthetic Division for Non-linear Polynomial Division," by Kim Gattis, Wichita State University.

"Running Races and Recursion," by Dan Fitzgerald, Kansas Newman College.

"Calculating Kendall's Tau Correlation Coefficient Using Weighted Data by SPSS," by Merillee Carlyle, Wichita State University.

"Combinatorial Investigations on 0-1 Matrices," by Dale Hignes, Wichita State University.

"Sober Spaces," by Dian Palenz, Wichita State University.

"On the Boundary Behavior of Non-parametric Minimal Surfaces in $R^3$," by Kirk Lancaster, Wichita State University.

Student Paper:

"An Application of Linear Interpolation in Two Variables for Brain Wave Displays," by Douglas P. Bogia, Washburn University.

Panel Discussion:

"Testing for Mathematics Proficiency," by John Carlson, Emporia State University (Moderator); Elton Beougher, Fort Hays State University; Rose Buschman, Garden City High School; Richard Driver, Topeka Public Schools; Ted Zerger, Kansas Wesleyan.

$X_1, X_2, \ldots, X_n$ of exchangeable r.v's. Let $T$ be a stopping-time defined on the sequence.

LEMMA 1. *If* $P(T \leqslant n-k)=1$, *the random variables* $(X_{T+1}, \ldots, X_{T+k})$ *and* $(X_1, \ldots, X_k)$ *have the same distribution.*

*Proof.* Let $A_t$ be the set of outcomes in $R_t$ that defines the event $T=t$; that is

$$\{T=t\} \Leftrightarrow \{(X_1, \ldots, X_t) \in A_t\}.$$

Letting $B$ be a set in $R_k$ we have

$$\begin{aligned}
&P[(X_{T+1}, \ldots, X_{T+k}) \in B] \\
&\qquad = \sum_{t=1}^{n-k} P[(X_1, \ldots, X_t) \in A_t \wedge (X_{t+1}, \ldots, X_{t+k}) \in B] \\
&\qquad = \sum_{t=1}^{n-k} P[(X_1, \ldots, X_t) \in A_t \wedge (X_{n-k+1}, \ldots, X_n) \in B] \\
&\qquad = P[(X_{n-k+1}, \ldots, X_n) \in B] = P[(X_1, \ldots, X_k) \in B].
\end{aligned}$$

The result in Example 2 follows from a corresponding lemma for an infinite sequence of exchangeable r.v.'s:

LEMMA 2. *For any* $k$, *the* r.v.'s $(X_{T+1}, \ldots, X_{T+k})$ *and* $(X_1, \ldots, X_k)$ *have the same distribution.*

We can express the results in Lemma 1 and Lemma 2 succinctly as follows: Exchangeable r.v's are invariant in distribution under stopping-time shifts.

Kallenberg [2] has proved the very interesting result that such invariance characterizes exchangeability; an infinite sequence of r.v.'s is exchangeable if and only if it is invariant under stopping-time shifts.

I am indebted to O. Kallenberg for valuable remarks.

## References

1. W. Feller, An Introduction to Probability Theory and Its Applications, vol. II, 2nd ed., Wiley, New York, 1971.
2. O. Kallenberg, Characterizations and embedding properties in exchangeability, Z. Wahrsch. Verw. Gebiete, 60 (1982) 249–281.
3. J. F. C. Kingman, Uses of exchangeability, Ann. Prob., 6 (1978) 183–197.

# SIMULTANEOUS COMPLEMENTS IN FINITE-DIMENSIONAL VECTOR SPACES

N. J. LORD
*Wolfson College, Oxford, England*

It is very well known that if $U$ is a subspace of the finite-dimensional vector space $V$, then $U$ has a complement in $V$; that is, there exists a subspace $X$ of $V$ such that the direct sum of $U$ and $X$, $U \oplus X$, is $V$. (See [1], p. 143.) Slightly less well known is that two subspaces $U$, $W$ of $V$, both of the same dimension, have a common complement. (A quick proof is by finite induction on the codimension of $U$; the inductive step resting on the observation that if $U \neq W$ then $U \cup W$ is never a subspace of $V$ ([1], p. 136). Choose $x \notin U \cup W$ and use the inductive hypothesis to obtain a common complement to span$(U, x)$ and span $(W, x)$.)

This note arose from the observation that there does not, in general, exist a common complement to three subspaces having a common dimension. Thus, for example, if $V$ is a 2-dimensional vector space over $\mathbb{Z}_2$, then $V$ has but three 1-dimensional subspaces; so it is

manifestly impossible to find a 1-dimensional subspace complementary to all three.

However, one is left with a feeling that this counterexample is somewhat "degenerate" in nature. (It does raise the interesting combinatorial problem: if $V$ is an $n$-dimensional vector space over a finite field, and if $1 \leqslant k \leqslant n-1$, what is the maximum number of distinct $k$-dimensional subspaces of $V$ that have a common complement?) Normally in elementary linear algebra one seeks fairly obvious counterexamples in familiar Euclidean spaces: the proof that one cannot find a counterexample in this situation turns out to be a pleasant exercise in what might be called the "generic method", blending topology and algebra. Indeed, we are able to prove rather more: in a finite-dimensional Euclidean space, given an arbitrary finite number (or even a countably infinite number) of subspaces all of the same dimension, there will exist a subspace that is a complement to every subspace in the given family.

For example, in 2-dimensions, this means that given any countable family of lines through the origin, there exists a line through the origin distinct from all of them. In this special case, an ad hoc argument can, of course, easily be supplied (the set of points on the unit circle which lie on the given lines is countable; the unit circle is not) which, in addition, suggests that if there exists one common complement there will, in some sense, exist many—a statement that will be made precise in the course of the proof. In 3-dimensions the result means that, given any countable family of lines through the origin, there exists a plane through the origin which contains none of them; and that given any countable family of planes through the origin, there exists a line through the origin which lies in none of the planes.

THEOREM. *Let $V$ be an $n$-dimensional vector space over the real numbers $\mathbb{R}$ (or the complex numbers), and let $\{U_i : i \in \mathbb{N}\}$ be a countable family of subspaces of $V$ each having the same dimension, $k$. Then there exists a subspace, $X$, of $V$ which is a complement in $V$ to all of the subspaces in the family.*

*Proof.* We write out the proof in the real case.

Let $(u_{1,i}, \ldots, u_{k,i})$ be a basis for $U_i$, let $V^{n-k}$ denote the $(n-k)$-fold Cartesian product of $V$ with itself, and define $f_i : V^{n-k} \to \mathbb{R}$ by:

$$f_i(v_1, \ldots, v_{n-k}) = \det(u_{1,i}, \ldots, u_{k,i}, v_1, \ldots, v_{n-k}).$$

Then $f_i$ is continuous (for any topology induced by a norm $\|\cdot\|$ on $V^{n-k}$), so that $O_i = f_i^{-1}(\mathbb{R} - \{0\})$ is an open subset of $V^{n-k}$. It is nonempty because there exists $(v_1, \ldots, v_{n-k})$ extending the basis for $U_i$ to a basis for $V$.

We claim that $O_i$ is dense in $V^{n-k}$. Let $\mathbf{v} = (v_1, \ldots, v_{n-k})$ be in $V^{n-k}$ and fix any $\mathbf{c}$ in $O_i$. Consider $p(x) = f_i(x\mathbf{v} + \mathbf{c})$ where $x \in \mathbb{R}$. Then $p$ is a polynomial function which is not identically zero since $p(0) = f_i(\mathbf{c}) \neq 0$ by the choice of $\mathbf{c}$. So there exists a positive $R$ such that $p(x) \neq 0$ for all $x \geqslant R$. Thus $x\mathbf{v} + \mathbf{c}$, and hence $\mathbf{v} + x^{-1}\mathbf{c}$, is in $O_i$ for all $x \geqslant R$. But

$$\|\mathbf{v} - (\mathbf{v} + x^{-1}\mathbf{c})\| = x^{-1}\|\mathbf{c}\| \to 0 \text{ as } x \to \infty,$$

which establishes the claim.

The Baire Category Theorem then implies that $\cap_{i=1}^{\infty} O_i$ is dense in $V^{n-k}$ and therefore nonempty. Choose any $(e_1, \ldots, e_{n-k})$ in $\cap_{i=1}^{\infty} O_i$, and, to complete the proof, let $X = \operatorname{span}(e_1, \ldots, e_{n-k})$.

Moreover, our preliminary suspicion that, if there were one common complement, there would be many, is confirmed in the strong sense that $\cap_{i=1}^{\infty} O_i$ (the set of possible choices) is dense in $V^{n-k}$.

**Reference**

1. I. N. Herstein, Topics in Algebra, Blaisdell, 1964.

## A SHORT PROOF OF CHEBYSHEV'S THEOREM*

N. COSTA PEREIRA
*Departamento de Fisica, Faculdade de Ciencias, Universidade de Lisboa, Lisbon, Portugal*

If we define $\psi(x)$, for $x > 0$, by $\psi(x) = \sum_{p^\alpha \leqslant x} \log p$ for all primes $p$ and positive integers $\alpha$, Chebyshev's classical theorem states that

$$(1) \qquad \liminf_{x\to\infty} \frac{\psi(x)}{x} \leqslant 1 \quad \text{and} \quad \limsup_{x\to\infty} \frac{\psi(x)}{x} \geqslant 1.$$

The prime number theorem is equivalent to the statement that $\lim_{x\to\infty}\psi(x)/x = 1$. Chebyshev's theorem is both weaker and easier to prove (see, e.g., Ingham [3], Hardy [2], or Ellison [1]). The purpose of this note is to give a direct and especially short proof of Chebyshev's theorem. We will use three elementary results:

$$(2) \qquad \log n! = \sum_{k\geqslant 1} \psi\left(\frac{n}{k}\right),$$

$$(3) \qquad \log n! = n\log n + O(n),$$

$$(4) \qquad \sum_{k=1}^{n} \frac{1}{k} = \log n + O(1).$$

Let $m$ be a fixed integer greater than one and $n$ an arbitrary integer greater than $m$. From (2) we get the double inequality:

$$(5) \qquad \sum_{k=1}^{[n/m]} \frac{1}{k}\frac{\psi\left(\frac{n}{k}\right)}{\frac{n}{k}} \leqslant \frac{\log n!}{n} \leqslant \sum_{k=1}^{[n/m]} \frac{1}{k}\frac{\psi\left(\frac{n}{k}\right)}{\frac{n}{k}} + \sum_{k=[n/m]+1}^{n} \frac{1}{k}\frac{\psi\left(\frac{n}{k}\right)}{\frac{n}{k}}.$$

If we define $L_m^-, L_m^+, \lambda_m$ by:

$$L_m^- = \inf_{x\geqslant m} \frac{\psi(x)}{x}, \quad L_m^+ = \sup_{x\geqslant m} \frac{\psi(x)}{x}, \quad \lambda_m = \sup_{1\leqslant x<m} \frac{\psi(x)}{x},$$

it follows at once from (5) that:

$$L_m^- \sum_{k=1}^{[n/m]} \frac{1}{k} \leqslant \frac{\log n!}{n} \leqslant L_m^+ \sum_{k=1}^{[n/m]} \frac{1}{k} + \lambda_m \sum_{k=[n/m]+1}^{n} \frac{1}{k}.$$

From identities (4) and (5) we then get

$$\begin{aligned} L_m^-(\log n - \log m + O(1)) &\leqslant \log n + O(1) \\ &\leqslant L_m^+(\log n - \log m + O(1)) + \lambda_m(\log m + O(1)). \end{aligned}$$

Hence we have

$$L_m^- \leqslant \frac{\log n + O(1)}{\log n - \log m + O(1)}$$

and

$$L_m^+ \geqslant \frac{\log n + O(1) - \lambda_m(\log m + O(1))}{\log n - \log m + O(1)}.$$

Since these inequalities are valid for every $n$ greater than $m$ and $\lambda_m$ is finite, we conclude that

*Supported in part by a grant from Secretaria de Estado da Cultura, Lisbon, Portugal.

$$L_m^- = \inf_{x \geqslant m} \frac{\psi(x)}{x} \leqslant 1 \quad \text{and} \quad L_m^+ = \sup_{x \geqslant m} \frac{\psi(x)}{x} \geqslant 1;$$

as $m$ approaches infinity we obtain (1).

**References**

1. W. J. Ellison, Les Nombres Premiers, Hermann, Paris, 1975.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th. ed., Oxford University Press, Oxford, 1962.
3. A. E. Ingham, The Distribution of Prime Numbers, Hafner, New York, 1965.

## PLANAR METRIC INEQUALITIES DERIVED FROM THE VANDERMONDE DETERMINANT

GENG-ZHE CHANG
*Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, The People's Republic of China*

This paper presents some new results that illustrate the power of complex numbers in planar geometry.

We begin with some notation.

Let $N$ denote the set of the first $n$ natural numbers. For $k = 1, 2, \dots, n-1$, let $S$ be a subset $(s_1, s_2, \dots, s_k)$ of $N$ such that $1 \leqslant s_1 < s_2 < \cdots < s_k \leqslant n$. We call $S$ a $k$-shuffle of $N$. The $(n-k)$-shuffle $(t_1, t_2, \dots, t_{n-k})$ is called the complementary shuffle of $S$ with respect to $N$ and is denoted by $S^-$ if

$$(t_1, t_2, \dots, t_{n-k}) \bigcup S = N.$$

After proving our main result, Theorem 1, we will give several geometric applications.

THEOREM 1. *If $z_1, z_2, \dots, z_n$ are distinct complex numbers, then for $k = 1, 2, \dots, n-1$, we have*

$$\sum_S \frac{(z_{t_1} z_{t_2} \cdots z_{t_{n-k}})^k}{\prod_{j=1}^{k} \prod_{l=1}^{n-k} (z_{t_l} - z_{s_j})} = 1, \tag{1}$$

*where the sum is extended over all $k$-shuffles $S = (s_1, s_2, \dots, s_k)$ and $S^- = (t_1, t_2, \dots, t_{n-k})$.*

*Proof.* We use the Vandermonde determinant

$$V(N) \equiv \begin{vmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ \vdots & \vdots & & \vdots \\ z_1^{n-1} & z_2^{n-1} & \cdots & z_n^{n-1} \end{vmatrix} = \prod_{\lambda < \mu} (z_\mu - z_\lambda). \tag{2}$$

If the Laplace expansion (see, for example, [6], p. 298) is applied to the first $k$ rows of $V(N)$, then we obtain

$$V(N) = \sum_S (-1)^{1+2+\cdots+k+|S|} (z_{t_1} z_{t_2} \cdots z_{t_{n-k}})^k \cdot V(S) V(S^-), \tag{3}$$

where the sum is extended over all $\binom{n}{k}$ of the $k$-shuffles $S$, $|S| = s_1 + s_2 + \cdots + s_k$, and $V(S)$ and $V(S^-)$ are the Vandermonde determinants determined by $z_{s_1}, z_{s_2}, \dots, z_{s_k}$ and $z_{t_1}, z_{t_2}, \dots, z_{t_{n-k}}$, respectively.

Note that $z_1, z_2, \dots, z_n$ are distinct. When $V(N)$ is replaced by the expression in the right-hand

side of (2), and $V(S)$ and $V(S^-)$ by similar expressions, we see clearly that

$$\frac{V(N)}{V(S)V(S^-)} = \sigma \prod_{j=1}^{k} \prod_{l=1}^{n-k} (z_{t_l} - z_{s_j}), \tag{4}$$

where

$$\sigma = \operatorname{sign} \prod_{j=1}^{k} \prod_{l=1}^{n-k} (t_l - s_j).$$

By an inversion of the sequence $(j_1, j_2, \ldots, j_n)$, a permutation of $N$, we mean a number pair $(j_p, j_q)$ with $p < q$ but $j_p > j_q$. Hence $\sigma = (-1)^{\tau(S,S^-)}$, where $\tau(S, S^-)$ designates the number of inversions of the sequence

$$(S, S^-) = (s_1, s_2, \ldots, s_k, t_1, t_2, \ldots, t_{n-k}).$$

In the sequence $(S, S^-)$, $(s_1 - 1)$ inversions are produced by $s_1$, $(s_2 - 2)$ inversions are produced by $s_2, \ldots, (s_k - k)$ inversions by $s_k$ and no other inversion exists; hence we have

$$\tau(S, S^-) = |S| - (1 + 2 + \cdots + k)$$

and (4) becomes

$$\frac{V(N)}{V(S)V(S^-)} = (-1)^{|S| - (1+2+\cdots+k)} \prod_{j=1}^{k} \prod_{l=1}^{n-k} (z_{t_l} - z_{s_j}). \tag{5}$$

Upon dividing both sides of (3) by $V(N)$ and using (5), we find that the identity (1) follows.
Theorems 2 and 3 show some of the geometric consequences of Theorem 1.

THEOREM 2. *Let $P_1, P_2, \ldots, P_n$ be $n$ distinct points in the Euclidean plane and let $P$ be some point in the same plane. Then for $k = 1, 2, \ldots, n-1$, we have*

$$\sum_S \frac{\left(|PP_{t_1}| \cdots |PP_{t_{n-k}}|\right)^k}{\prod_{j=1}^{k} \prod_{l=1}^{n-k} |P_{s_j} P_{t_l}|} \geqslant 1, \tag{6}$$

*where the sum is extended over all $k$-shuffles $S = (s_1, s_2, \ldots, s_k)$, $|P_{s_j} P_{t_l}|$ denotes the distance between points $P_{s_j}$ and $P_{t_l}$, and $S^- = (t_1, t_2, \ldots, t_{n-k})$.*

*Proof.* Set a complex coordinate system on the plane with point $P$ as the origin. Assume that $P_1, P_2, \ldots, P_n$ have complex coordinates $z_1, z_2, \ldots, z_n$, respectively. Upon taking the modulus of each term in the left-hand side of (1), we obtain

$$\sum_S \frac{\left(|z_{t_1}| \cdots |z_{t_{n-k}}|\right)^k}{\prod_{j=1}^{k} \prod_{l=1}^{n-k} |z_{t_l} - z_{s_j}|} \geqslant 1, \tag{7}$$

which is equivalent to (6) in the present case.
Equality in (6) may occur, as will be shown in Theorem 3.

REMARK. Putting $k = 1$ into (6) we have

$$\sum_{i=1}^{n} \prod_{j \neq i} \frac{|PP_j|}{|P_i P_j|} \geqslant 1. \tag{8}$$

Inequality (8) establishes the Euclidean conjecture of K. B. Stolarsky [10] in the case $n = 3$; for $n \geqslant 4$ this conjecture is still open. In fact, inequality (8) for 4 points was originally established by T. Hayashi [5] in 1913. In 1975 M. S. Klamkin proved it via inversion and the use (see [8]) of polar moments of inertia; he later discussed Hayashi's method and developed it further in [9]. Whether

a "one line proof" of Hayashi's result can be obtained by rearranging the Cayley-Menger determinant remains open: see the discussion in [14], p. 233.

THEOREM 3. *Let $P_1, P_2, \ldots, P_n$ denote $n$ distinct points on a circle of unit radius. For each $k$-shuffle $S = (s_1, s_2, \ldots, s_k)$ with its complement $S^- = (t_1, t_2, \ldots, t_{n-k})$, define*

$$d(S) \equiv \prod_{j=1}^{k} \prod_{l=1}^{n-k} |P_{s_j} P_{t_l}|. \tag{9}$$

*Then*

$$\sum_S \frac{1}{d(S)} \geqslant 1, \tag{10}$$

*where the sum is extended over all $k$-shuffles of $N$. Equality is attained if and only if the points form a regular $n$-gon.*

*Proof.* Inequality (10) follows directly from (6) if we put the point $P$ at the center of the circle. It remains to consider under what conditions equality holds in (10).

Suppose that $P_1, P_2, \ldots, P_n$ form a regular $n$-gon. Without loss of generality we may assume that the complex coordinate of $P_m$ is $z_m = w^m$, $m = 1, 2, \ldots, n$, where

$$w = \exp(i2\pi/n) = \cos(2\pi/n) + i\sin(2\pi/n).$$

We are going to show that

$$\prod_{j=1}^{k} \prod_{l=1}^{n-k} \left(1 - z_{s_j} \bar{z}_{t_l}\right) \tag{11}$$

or equivalently

$$\prod_{j=1}^{k} \prod_{l=1}^{n-k} \left(1 - w^{s_j - t_l}\right) \tag{12}$$

is a positive number.

Since we have

$$\left[\prod_{l=1}^{n-k} \left(1 - w^{s_j - t_l}\right)\right] \prod_{m \neq j} \left(1 - w^{s_j - s_m}\right) = (1 - w)(1 - w^2) \cdots (1 - w^{n-1}) = n,$$

multiplication of (12) and

$$\prod_{j=1}^{k} \prod_{m \neq j} \left(1 - w^{s_j - s_m}\right) \tag{13}$$

yields $n^k$, a positive number. It is obvious that (13) is equal to

$$\prod_{1 \leqslant m \neq j \leqslant k} \left(1 - w^{s_j - s_m}\right) = \prod_{1 \leqslant j < m \leqslant k} \left|\left(1 - w^{s_j - s_m}\right)\right|^2 > 0.$$

It follows that (11) is positive and so is each term in the left-hand side of (1). The positivity of each term yields equality in (7) and also in (10).

Conversely, assume that equality holds in (10) and hence in (7). It is necessary that each term in the left-hand side of (1) is a positive real number. Set $z_j = \exp(i\theta_j)$, $j = 1, 2, \ldots, n$. We must have for each $k$-shuffle $S = (s_1, s_2, \ldots, s_k)$ of $N$ that

$$\prod_{j=1}^{k} \prod_{l=1}^{n-k} \left\{1 - \exp\left[i\left(\theta_{s_j} - \theta_{t_l}\right)\right]\right\} \tag{14}$$

is a real number. Since

$$1-\exp\left[i\left(\theta_{s_j}-\theta_{t_l}\right)\right]=-2\sin\left[\left(\theta_{s_j}-\theta_{t_l}\right)/2\right]\exp\left[i\left(\pi+\theta_{s_j}-\theta_{t_l}\right)/2\right],$$

(14) becomes

$$(-2)^{k(n-k)}\left[\prod_{j=1}^{k}\prod_{l=1}^{n-k}\sin\left(\theta_{s_j}-\theta_{t_l}\right)/2\right]\cdot\exp\left\{i\left[k(n-k)\pi+\sum_{j,l}\left(\theta_{s_j}-\theta_{t_l}\right)\right]\Big/2\right\}.$$

Since the above quantity is real, we have

$$k(n-k)\pi+\sum_{j,l}\left(\theta_{s_j}-\theta_{t_l}\right)=2m\pi, \tag{15}$$

where $m$ is an integer. Now

$$\begin{aligned}\sum_{j,l}\left(\theta_{s_j}-\theta_{t_l}\right)&=(n-k)\sum_{j=1}^{k}\theta_{s_j}-k\sum_{l=1}^{n-k}\theta_{t_l}\\&=n\sum_{j=1}^{k}\theta_{s_j}-k\sum_{j=1}^{n}\theta_j,\end{aligned}$$

so it follows from (15) that for any two $k$-shuffles $S$ and $S'$

$$\sum_{j=1}^{k}\theta_{s_j}-\sum_{j=1}^{k}\theta_{s'_j}=2m\pi/n. \tag{16}$$

Since $z_1, z_2, \ldots, z_n$ are distinct, we conclude by (16) that $z_1, z_2, \ldots, z_n$, or equivalently $P_1, P_2, \ldots, P_n$ are vertices of a regular $n$-gon. This proves the result.

In [15], a result due to H. S. Shapiro was stated without proof as "let $P_1, P_2, \ldots, P_n$ denote $n$ points on a circle of unit radius, and set $d_k$ equal to the product of the distance from $P_k$ to the other points. Then

$$\sum_{k=1}^{n}\frac{1}{d_k}\geqslant 1,$$

equality being attained if and only if the points form a regular $n$-gon." Clearly Theorem 3 of the present paper generalizes Shapiro's statement.

For a numerical example take $n=5$ and let $a$ and $d$ denote the side and the diagonal of the pentagon, respectively.

For $k=1$, by Theorem 3 and also by Shapiro's statement, we have

$$a^2d^2=5.$$

For $k=2$, by Theorem 3, we get

$$5(a^2+d^2)=(ad)^4.$$

REMARK. There are many easy to state but open problems on inequalities among the distances determined by $n$ points in $d$-dimensional Euclidean space. The work of J. B. Kelly on hypermetric spaces sheds some light on these problems; see [7] and the references on p. 31. In many of these problems (as in Theorem 3 of this paper) the points are constrained to lie on the boundary of a disc or surface of a sphere; see Alexander and Stolarsky [2], Stolarsky [11], [13], and Harmon [4]. Recently some remarkable progress has been made by J. Beck [3]. For an inequality involving points near the unit circle (and involving an expression reminiscent of (11)) see Alexander [1].

## References

1. J. R. Alexander, On an inequality of J. W. S. Cassels, this MONTHLY, 79 (1972) 883–884.

2. J. R. Alexander and K. B. Stolarsky, Extremal problems of distance geometry related to energy integrals, Trans. Amer. Math. Soc., 193 (1974) 1–31.

3. J. Beck, On the sum of distances between $n$ points on a sphere—an application of the theory of irregularities of distribution to discrete geometry, in preparation.

4. G. Harmon, Sums of distances between points of a sphere, Internat. J. of Math. and Math. Sci., 5 (1982) 707–714.

5. T. Hayashi, Two theorems on complex numbers, Tôhoku Math. J., 4 (1913/14) 68–70.

6. F. E. Hohn, Elementary Matrix Algebra, 3rd ed., Macmillan, New York, 1973.

7. J. B. Kelly, Hypermetric spaces, in The Geometry of Metric and Linear Spaces, L. M. Kelly, ed., Lecture Notes in Math. 490, Springer-Verlag, Berlin, 1975, pp. 17–31.

8. M. S. Klamkin, Geometric inequalities via the polar moment of inertia, Math. Mag., 48 (1975) 44–46.

9. ______, Triangle inequalities from the triangle inequality, Elem. der Math., 34 (1979) 49–55.

10. K. B. Stolarsky, Cubic triangle inequalities, this MONTHLY, 78 (1971) 879–881.

11. ______, Sums of distances between points on a sphere II, Proc. Amer. Math. Soc., 41 (1973) 575–582.

12. ______, The sum of the distances to $N$ points on a sphere, Pacific J. Math., 57 (1975) 563–573.

13. ______, The sum of the distances to certain pointsets on the unit circle, Pacific J. Math., 59 (1975) 241–251.

14. ______, Problem 1, in The Geometry of Metric and Linear Spaces, L. M. Kelly, ed., Lecture Notes in Math. 490, Springer-Verlag, 1975, p. 233.

15. Editor's comments for problem E 1023, this MONTHLY, 60 (1953) 119–121.

## COMBINATORIAL PROOFS OF PASCAL'S FORMULA FOR SUMS OF POWERS OF THE INTEGERS

JEROME L. PAUL
*Department of Computer Science, University of Cincinnati, Cincinnati, OH 45221*

Pascal's formula (cf. [1]) involving the sum of the $k$th powers of the $n$ successive terms $a, a+d, \dots, a+(n-1)d$ in an arithmetic progression can be written:

$$\sum_{k=0}^{r-1}\sum_{j=1}^{n}\binom{r}{k}d^{r-k}(a+(j-1)d)^k=(a+nd)^r-a^r. \tag{1}$$

The formula is proved by interchanging the order of summation, using the binomial theorem, and noting the resulting telescoping sum (see e.g. [3]). The following related formula can be established in a similar manner:

$$\sum_{k=0}^{r-1}\sum_{j=1}^{n}(-1)^{r-k+1}\binom{r}{k}d^{r-k}(a+jd)^k=(a+nd)^r-a^r. \tag{2}$$

In [2] a combinatorial proof of (1) was given for the special case $a=d=1$. This case is of particular interest, since the formula then involves

$$1^k+2^k+\cdots+n^k=S_n(k).$$

The proof consisted in enumerating the points in a hypercube of integer lattice points in $r$-dimensional space via a partition of the cube into $r$ disjoint subsets. In this note a combinatorial proof of (2) will be given for the case $a=0$, $d=1$, a case which also involves the sum $S_n(k)$. The proof will again consist in enumerating the points in a hypercube of lattice points using $r$ subsets, but here the subsets will not be disjoint, allowing a direct application of the inclusion-exclusion principle. Recall that the inclusion-exclusion principle is a generalization to $r$ sets of the familiar formula

$$|A\cup B|=|A|+|B|-|A\cap B|,$$

where $|X|$ denotes the number of points in the set $X$. The principle is stated as formula (4), and can be established by a straightforward counting argument (see e.g. [4]).

Geometric motivation for the particular choice of subsets is readily obtained in dimensions $r=2, 3$, and the reader is urged to draw the appropriate pictures. A description of how these proofs can be extended to (1) and (2) for the case of arbitrary nonnegative integers $a$ and $d$ will

also be given. Throughout the following, let $C^r(m)$ denote the hypercube of integer lattice points in $r$-dimensional space having $m$ points on a side,

$$C^r(m) = \{(x_1, \ldots, x_r) \in Z^r: \quad 0 \leqslant x_i \leqslant m-1, 1 \leqslant i \leqslant r\},$$

so that $|C^r(m)| = m^r$, where $Z$ denotes the integers.

Setting $a = 0, d = 1$ into (2), replacing $k$ by $r-k$, and using $\binom{r}{k} = \binom{r}{r-k}$, this special case will be proved in the equivalent form

$$\sum_{k=1}^{r} (-1)^{k+1}\binom{r}{k}(1^{r-k} + 2^{r-k} + \cdots + n^{r-k}) = n^r. \tag{3}$$

For a fixed $i, 1 \leqslant i \leqslant r$, let $K_i$ be the subset of $C^r(n)$ consisting of those points $(x_1, \ldots, x_r)$ for which $x_i \geqslant x_j, 1 \leqslant j \leqslant r$. The inclusion-exclusion principle applied to the sets $K_i$ yields

$$\left|\bigcup_{i=1}^{r} K_i\right| = \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant r} (-1)^{k+1}|K_{i_1} \cap \cdots \cap K_{i_k}|, \tag{4}$$

where the summation is over all $k$-combinations of integers in $\{1, \ldots, r\}, k = 1, \ldots, r$. Note that for any choice of $k$ indices $1 \leqslant i_1 < \cdots < i_k \leqslant r$,

$$|K_{i_1} \cap \cdots \cap K_{i_k}| = \sum_{j=1}^{n} j^{r-k},$$

since if $x_{i_1} = \cdots = x_{i_k} = j-1$, then the remaining $r-k$ coordinates can have any of the $j$ values between 0 and $j-1$. Since there are $\binom{r}{k}$ ways to choose the $k$ indices $1 \leqslant i_1 < \cdots < i_k \leqslant r$, and

$$\left|\bigcup_{i=1}^{r} K_i\right| = |C^r(n)| = n^r,$$

note that (3) follows from (4).

The proof above is easily generalized to establish (2) in the case of arbitrary nonnegative integers $a$ and $d$. For $a = 0, d > 0$, the generalization amounts to an "expansion" of each point of $C^r(n)$ into an equivalence class of points in $C^r(nd)$, and when $a > 0, C^r(1+n)$ is "expanded" to $C^r(a+nd)$. To see this, let $\phi$ be a step function defined on the nonnegative integers by

$$\phi(x) = \begin{cases} -1, & 0 \leqslant x \leqslant a-1, \\ j, & a + jd \leqslant x \leqslant a + (j+1)d - 1, j \geqslant 0. \end{cases}$$

Equivalently, $\phi(x) = \max\left\{\left[\frac{x-a}{d}\right], -1\right\}$, where $[y]$ denotes the greatest integer $\leqslant y$. For a fixed $i, 1 \leqslant i \leqslant r$, let $K_i$ be the subset of $C^r(a+nd)$ consisting of those points $(x_1, \ldots, x_r)$ for which $\phi(x_i) \geqslant \phi(x_j)$, $1 \leqslant j \leqslant r$. Now for any choice of $k$ indices $1 \leqslant i_1 < \cdots < i_k \leqslant r$, it is easily verified in a manner analogous to that above that

$$|K_{i_1} \cap \cdots \cap K_{i_k}| = a^r + \sum_{j=1}^{n} d^k(a+jd)^{r-k}.$$

An application of (4) to the sets $K_i$ then yields

$$\sum_{k=1}^{r} (-1)^{k+1}\binom{r}{k}\left(a^r + \sum_{j=1}^{n} d^k(a+jd)^{r-k}\right) = (a+nd)^r, \tag{5}$$

which is equivalent to (2). (Note that $\sum_{k=1}^{r}(-1)^{k+1}\binom{r}{k} = 1$).

With the aid of the function $\phi$, the proof of (1) given in [2] when $a = d = 1$ can be generalized

in a similar manner to establish (1) for arbitrary nonnegative integers $a$ and $d$. The details are left to the reader.


### References

1. C. B. Boyer, Pascal's formula for sums of powers of the integers, Scripta Mathematica, 9 (1943) 237–244.
2. J. L. Paul, On the sum of the $k$th powers of the first $n$ integers, this MONTHLY, 78 (1971) 271.
3. B. Turner, Sums of powers of integers via the binomial theorem, Math. Mag., 53 (1980) 92–96.
4. D. E. Knuth, The Art of Computer Programming, vol. 1, 2nd ed., Addison-Wesley, Reading, MA, 1973, p. 178.


# THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## A NOTE ON COMPLEX ITERATION


I. N. BAKER
*Department of Mathematics, Imperial College, London, England*

P. J. RIPPON
*Faculty of Mathematics, The Open University, Milton Keynes, England*


The aim of this note is to describe how a recent result about the sequence

$$a, a^a, a^{(a^a)}, \ldots, \qquad a \in \mathbb{C}, \tag{1}$$

might be included in a course on complex analysis (one that is based for instance on [1]). The study of the sequence (1) goes back to Euler [5] who showed that, for $a > 0$, it is convergent if and only if

$$e^{-e} \leqslant a \leqslant e^{1/e}.$$

To discuss the general case it is convenient to let $c$ be some fixed determination of $\log a$ and to put

$$f(z) = e^{cz}, \qquad z \in \mathbb{C}, \tag{2}$$

so that, if $f^{n+1} = f \circ f^n$, $n = 1, 2, \ldots,$ with $f^1 = f$, then

$$f^{n+1}(0), \qquad n = 1, 2, \ldots, \tag{3}$$

is a well-defined version of (1).

It was shown by Carlsson [4] that if $f^n(0) \to w$ as $n \to \infty$ and $f^n(0) \neq w$, $n = 1, 2, \ldots,$ then $c$ must lie in the closure of a certain cardioid

$$D = \{ te^{-t} : |t| < 1 \},$$

which is illustrated in Fig. 1.

To prove Carlsson's result, note that $\lim_{n\to\infty} f^n(0) = w$ implies that $w = f(w) = e^{cw}$. Thus if $t = cw$, then $w = e^t$ and so $c = te^{-t}$. To prove that $|t| \leqslant 1$ we use the fact that $f'(w) = ce^{cw} = t$.

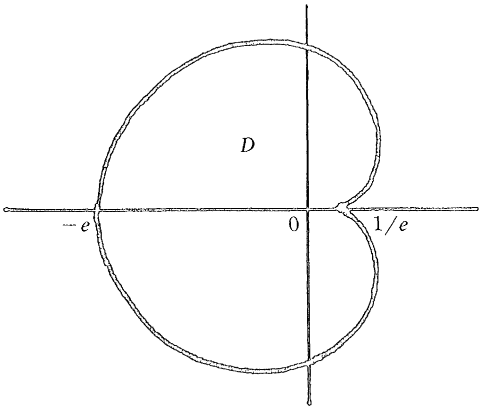



FIG. 1

Since $f^n(0) \neq w$, $n = 1, 2, \ldots$, we have

$$\lim_{n \to \infty} \left( \frac{f^{n+1}(0) - w}{f^n(0) - w} \right) = \lim_{n \to \infty} \left( \frac{f(f^n(0)) - f(w)}{f^n(0) - w} \right) = f'(w) = t. \tag{4}$$

This implies that $|t| \leqslant 1$, as required. Also, if $t$ is not real, then, according to (4), the convergence of $f^n(0)$ to $w$ is eventually spiral-like. Some illustrations of the kinds of convergence which occur are given in [8].

We shall now give a sufficient condition for the convergence of (3). Previous results in this direction are surveyed in [7].

THEOREM. *The sequence* (3) *is convergent whenever $c$ lies in $D$.*

This result was proved in [2] using the theory of iteration due to Fatou and Julia (see, for instance, [6]). We prove it below using only those parts of this general theory which seem to be essential. Note that the case $c = 0$ can be disposed of immediately.

If $c = te^{-t}$ with $0 < |t| < 1$, then $w = e^t$ is a fixed point of $f$ with $f'(w) = t$. Thus

$$\lim_{z \to w} \left( \frac{f(z) - w}{z - w} \right) = \lim_{z \to w} \left( \frac{f(z) - f(w)}{z - w} \right) = f'(w) = t,$$

and since $|t| < 1$ there exist $r > 0$ and $\lambda < 1$ such that

$$\left| \frac{f(z) - w}{z - w} \right| \leqslant \lambda, \qquad 0 < |z - w| < r.$$

We deduce that $f^n(z) \to w$ uniformly as $n \to \infty$ for $|z - w| < r$.

Now consider the set $\Omega$ of those complex numbers in some neighbourhood of which the sequence $f^n \to w$ uniformly as $n \to \infty$. To prove the theorem we show that $0 \in \Omega$. It is clear that $\Omega$ is open, that $w \in \Omega$ and that $\Omega$ is completely invariant under $f$, by which we mean that

$$z \in \Omega \Leftrightarrow f(z) \in \Omega.$$

Also, by the Heine-Borel theorem, the sequence $f^n \to w$ as $n \to \infty$ uniformly on any compact subset of $\Omega$. It follows from this that each component of $\Omega$ is simply connected. Indeed if $\Gamma$ is any Jordan curve in $\Omega$, then $f^n - w \to 0$ uniformly on $\Gamma$ and so $f^n - w \to 0$ uniformly in the interior of $\Gamma$, by the maximum principle. This implies that $\Omega$ has no "holes".

Let $\Omega_w$ be the component of $\Omega$ which contains $w$ and suppose that $0 \notin \Omega_w$. We shall show that this leads to a contradiction. Since $\Omega_w$ is simply connected, we can choose in $\Omega_w$ [1, p. 143] a single-valued analytic branch $g$ of

$$f^{-1}(z) = (\log z)/c,$$

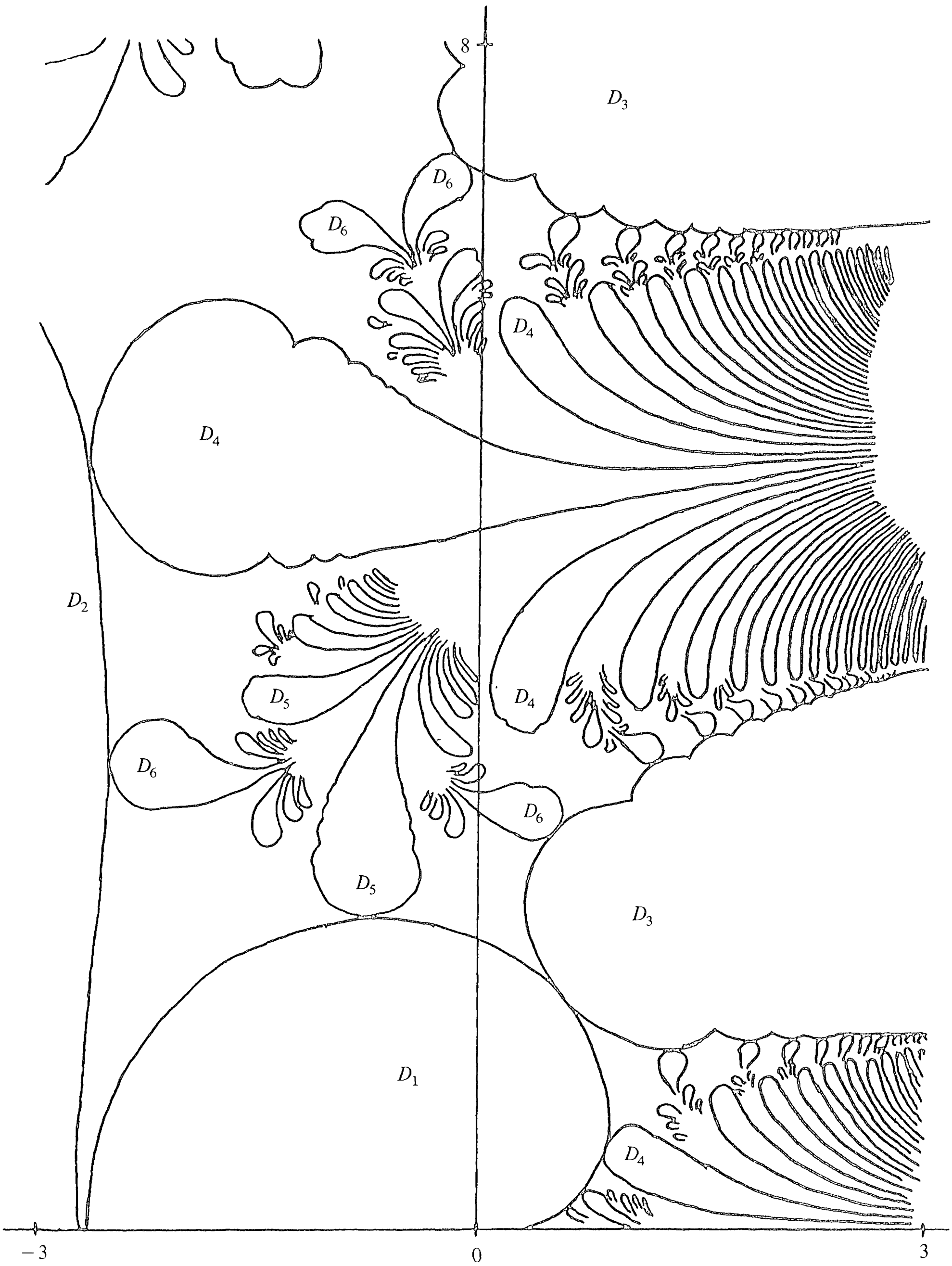



FIG. 2

such that $g(w) = w$. The complete invariance of $\Omega$ and the continuity of $g$ then imply that $g(\Omega_w) \subseteq \Omega_w$ and we deduce, from Schwarz's lemma, that

$$|g'(w)| \leqslant 1.$$

Thus $|f'(w)| = |g'(w)|^{-1} \geq 1$ and this contradiction completes the proof.

The form of Schwarz's lemma used here is as follows.

LEMMA. *Let $g$ be an analytic function which maps a simply connected domain $\Omega_w \neq \mathbb{C}$ into itself with fixed point $w \in \Omega_w$. Then*

$$|g'(w)| \leqslant 1.$$

This can be proved by using the Riemann mapping theorem [**1**, p. 222] to choose a conformal mapping $\phi$ from $\Omega_w$ onto the unit disc $U$, such that $\phi(w) = 0$. The result follows by applying the usual form of Schwarz's lemma to $\phi \circ g \circ \phi^{-1}$, which maps $U$ to $U$ and fixes 0.

If a more elementary argument is required, then we can choose, as in the first stage of the proof of the Riemann mapping theorem [**1**, p. 222], an analytic function $\phi$ from $\Omega_w$ into $U$ such that $\phi'(w) \neq 0$. The sequence of functions $\phi_n = \phi \circ g^n$, $n = 1, 2, \ldots$, is then uniformly bounded in $\Omega_w$ and so, by [**1**, p. 122], the sequence $\phi'_n$ is uniformly bounded in some neighbourhood of $w$. But

$$\phi'_n(w) = \phi'(w)(g'(w))^n,$$

since $g(w) = w$ and, in view of the fact that $\phi'(w) \neq 0$, we deduce that $|g'(w)| \leqslant 1$.

REMARKS. In [**2**] we also describe the behaviour of the sequence (3) for most boundary points of $D$. It is convergent if $c = te^{-t}$, where $t$ is an $n$th root of unity, but divergent almost everywhere on $\partial D$. The behaviour of (3) for values of $c$ outside $\overline{D}$ is described in a forthcoming paper [**3**]. Briefly, in the $c$-plane there are infinitely many disjoint domains, in each of which the sequence (3) forms a number of convergent subsequences. In those domains illustrated in Fig. 2 the label $D_k$ indicates that, for the corresponding numbers $c$, the sequence (3) has $k$ convergent subsequences $\{f^{nk+j}(0)\}_{n=1}^{\infty}$, $j = 1, 2, \ldots, k$, each with a distinct limit.

The domain $D_1$ is the cardioid $D$ discussed above. Each of the domains $D_k$, $k \geqslant 2$, is simply connected and unbounded, there is a single $D_2$, but for $k \geqslant 3$ there appear to be infinitely many $D_k$. The relative positions of these domains are described in more detail in [**3**], but there remain a number of open questions. For example, we do not know whether the union of all these domains $D_k$ is dense in $\mathbb{C}$.

### References


**1**. L. V. Ahlfors, Complex Analysis, McGraw-Hill, 1966.

**2**. I. N. Baker and P. J. Rippon, Convergence of infinite exponentials, Ann. Acad. Sci. Fenn. Ser. AI Math., 8 (1983) 179–186.

**3**. ______, Iteration of exponential functions, Ann. Acad. Sci. Fenn. Ser. A I Math. (to appear).

**4**. A. Carlsson, Om Iterade Functioner, Ph.D. Thesis, Uppsala, 1907.

**5**. L. Euler, De formulis exponentialibus replicatus, Opera Omnia, Series Primus XV, 268–297; Acta Acad. Petropolitanae, 1 (1777) 38–60.

**6**. P. Fatou, Sur l'itération des fonctions transcendantes entières, Acta Math., 47 (1926) 337–370.

**7**. R. A. Knoebel, Exponentials reiterated, this MONTHLY, 88 (1981) 235–252.

**8**. P. J. Rippon, Infinite exponentials, Math. Gaz., 67 (1983) 189–196.


## COMPACTNESS AND CLOSEDNESS IN LOCALLY COMPACT HAUSDORFF SPACES


WARREN PAGE

*Department of Mathematics, New York City Technical College, Brooklyn, NY 11201*


The purpose of this note is to illuminate a reasoning error in point set topology that is seductively easy for experts as well as students to make—namely, the belief that

*If $Y$ is a locally compact subset of a locally compact Hausdorff space $X$, then $Y \cap K$ is compact for every compact set $K \subset X$.*

(A simple counterexample is the open segment $(0, 1)$, which is a locally compact subset of the real line, although its intersection with the compact interval $[0, 1]$ is not compact.) Since this unjustified assumption has, in fact, led to circular reasoning in a number of sources in proving the

Pontryagin Duality Theorem (e.g., [**1**], [**3**]), it should be instructive to illustrate a few contexts in which the above conclusion is and is not warranted. We hope that this note, by calling attention to these deficient proofs, will also be of value to all who continue to use and refer to Naimark's and Rudin's classical texts.

It will be useful to recall the following characterization of closedness in locally compact Hausdorff ($\mathrm{LCT}_2$) spaces.

THEOREM 1. *A subset $Y$ of an* $\mathrm{LCT}_2$ *space $X$ is closed in $X$ if and only if $Y \cap K$ is compact for every compact subset $K \subset X$.*

*Proof.* If $x_0 \in \overline{Y} - Y$ and $u$ is a relatively compact neighborhood of $x_0 \in X$, then

$$x_0 \in \left(\overline{Y \cap \bar{u}}\right) - (Y \cap \bar{u}).$$

Thus, $Y \cap \bar{u}$ is not closed and therefore not compact in $X$. Assume, conversely, that $Y$ is closed in $X$ and let $K$ be compact in $X$. If

$$Y \cap K \subset \bigcup_A (Y \cap O_a)$$

for $X$-open sets $\{O_a : a \in A\}$, then

$$Y \cap K \subset \bigcup_A (K \cap O_a).$$

Since $Y \cap K$ is compact in $K$, we have

$$Y \cap K \subset \bigcup_B (K \cap O_a)$$

for some finite subset $B \subset A$. In particular,

$$Y \cap K \subset \bigcup_B (Y \cap O_a)$$

and $Y \cap K$ is seen to be compact in $X$. □

Now consider the Pontryagin Duality Theorem for locally compact Hausdorff abelian ($\mathrm{LCT}_2\mathrm{A}$) groups. If $X$ is the second dual group (i.e., the dual group of $G'$) associated with an $\mathrm{LCT}_2\mathrm{A}$ group $G$, the natural embedding $\delta$: $G \to X$ with $\{\delta(g)\}(\chi) = \chi(g)$ for every $\chi \in G'$ is an into topological group isomorphism. (See, for instance, [**1**], [**2**], [**3**].) What remains is to establish that the $\mathrm{LCT}_2\mathrm{A}$ subgroup $Y = \delta(G)$ is all of the $\mathrm{LCT}_2\mathrm{A}$ group $X$. One approach is to first prove that $Y$ is closed in $X$ and then show (independently of this) that $Y$ is dense in $X$. It is in this first step where several noted texts go wrong.

EXAMPLE 1 ([**3**], p. 29). Suppose $x_0 \in \overline{Y}$ and let $u$ be an $X$-relatively compact neighborhood of $x_0$. *Since $Y$ is* $\mathrm{LCT}_2$, *the set $Y \cap \bar{u}$ is compact in $X$.* Therefore, $Y \cap \bar{u}$ is closed in $X$, and

$$x_0 \in \overline{Y \cap \bar{u}} = Y \cap \bar{u} \subset Y.$$

Thus, $Y$ is closed in $X$. But this italicized assertion is unjustified, and so this proof is untenable.

EXAMPLE 2 ([**1**], p. 418). Let $Y_\infty$ and $X_\infty$ be the one-point compactifications of $Y$ and $X$, respectively. *Then, $Y_\infty$ is also a topological subspace of $X_\infty$.* Since $Y_\infty$ is closed in $X_\infty$, it follows that $Y$ is closed in $X$. But, although every $Y$-compact set is $X$-compact, $Y \cap K$ cannot be assumed compact in $Y$ for an $X$-compact set $K$. Thus, although the topology of $Y_\infty$ is weaker than the $X_\infty$-induced topology, the reverse inclusion has not been established. The italicized assertion is therefore unjustified.

It is interesting to note how the above deficiencies can be overcome, and the proofs repaired, by reordering the final two steps in these erroneous proofs. Implicit in this approach (or any other proof of the Pontryagin Duality Theorem) is the realization that topological arguments alone,

without using the group's structure, are not sufficient for proving $Y$ closed in $X$. The proof that $Y = \delta(G)$ is closed in $X$ is simple once it has first been established that $Y$ is dense in $X$.

THEOREM 2. *If $Y$ is a dense subgroup of a $T_2$ group $X$ and if $Y$ is* LC *in the topology induced by $X$, then $Y = X$.*

*Proof.* For any $y_0 \in Y$, there is an $X$-neighborhood $v(y_0)$ such that the $Y$-closure $Z$ $(= Y \cap \overline{Y \cap v})$ of $Y \cap v$ is compact in $Y$, hence compact in $X$, and therefore closed in $X$. Now suppose $p \in v$, and let $u$ be a neighborhood of $p$. Then $u \cap v$ is a nonempty neighborhood which, because $Y$ is dense in $X$, must intersect $Y$. Thus, $u$ intersects $Y \cap v \subset Z$. Since $Z$ is closed in $X$, it follows that $p \in Z$. Thus, $v \subset Z$ and hence $v \subset Y$. Finally, if $x \in X = \overline{Y}$, the neighborhood $vx^{-1}$ contains some $y \in Y$. Therefore, $x \in y^{-1}v$. Since $Y$ is a group and $v \subset Y$, it follows that $x \in Y$. □

Note (by way of contrast with Example 1) that when $\overline{Y} = X$, we are assured that $Y \cap \bar{u}$ is compact. This, of course, can also be proved without requiring that $Y$ be dense in $X$. Indeed,

THEOREM 4. *An* $\mathrm{LCT}_2$ *subgroup $Y$ of an* $\mathrm{LCT}_2$ *group $X$ has the following equivalent properties*:
(i) *$Y$ is closed in $X$,*
(ii) *$Y \cap K$ is $Y$-compact for every compact set $K \subset X$,*
(iii) *$Y_\infty$ is a topological subspace of $X_\infty$.*

*Proof.* An $\mathrm{LCT}_2$ subgroup $Y$ is complete and therefore closed in $X$. Thus, (i) and (Theorem 1) its equivalent (ii) hold. Moreover, (ii) ⇒ (iii) ⇒ (i) based on the type of arguments used in Example 2.

The author wishes to thank George Bachman for calling his attention to the incorrect proofs in the literature.

**References**

1. M. Naimark, Normed Rings, Noordhoff, Groningen, 1959.
2. W. Page, Topological Uniform Structures, Wiley-Interscience, New York, 1978.
3. W. Rudin, Fourier Analysis on Groups, Wiley, New York, 1962.

---

# PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1986. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

**Correction to E 3090** (May 1985 issue, p. 359). The first part of the problem should read:

Let $f:[0,+\infty)\to(0,\infty)$ be a continuous function. Let $h>0$ be given and let $g:[0,+\infty)\to R$ be defined by

$$g(x)=x\left(\frac{f(x)}{f(x+h)}-1\right).$$

(The remainder of the problem is as stated.)

E 3099. *Proposed by Weixuan Li and Edward T. H. Wang, Wilfrid Laurier University, Canada.*

Let $a_1\leqslant a_2\leqslant\cdots\leqslant a_n$ be $n$ nonnegative reals ($n\geqslant 2$) such that

$$\sum_{i=1}^{n}a_ia_{i+1}=1\qquad(a_{n+1}=a_1).$$

Determine the minimum value of $\sum_{i=1}^{n}a_i$.

E 3100. *Proposed by C. J. Smyth, James Cook University, Townsville, Australia.*

Given a rational number $p/q$, show that there are conjugate algebraic numbers, $\alpha,\alpha'$, with $|\alpha|\neq 0,1$ and $|\alpha|^{p/q}=|\alpha'|$.

E 3101. *Proposed by Charles Vanden Eynden, Illinois State University.*

Let $d(n)$ be the number of positive divisors of $n>0$. Show that the number of solutions $x$ to $d(nx)=n$ is 1 if $n=4$, $t!$ if $n$ is the product of $t$ distinct primes ($t\geqslant 0$), and infinite otherwise.

E 3102. *Proposed by J. P. Lambert, University of Alaska.*

Prove or disprove: For all integers $m,n\geqslant 1$,

$$\frac{\sum_{j=1}^{m}(2j-1)^{2n}}{\sum_{j=1}^{m}(2j)^{2n}}<\left(\frac{2m}{2m+1}\right)^{2n+1}.$$

E 3103. *Proposed by Douglas B. Tyler, Fullerton, California.*

Show that

$$\sum_{k=1}^{\infty}\frac{\zeta(2k)}{k(2k+1)\cdot 2^{2k}}=\ln\pi-1,$$

where $\zeta$ is the Riemann zeta function.

E 3104. *Proposed by Jay Hook, Florida International University.*

Let $A$ be a closed and bounded subset of $\mathbb{R}$ of Lebesgue measure 1.

(a) Show that $A$ contains two points an integer distance apart.
(b) Show that (a) becomes false if "and bounded" is omitted in the definition of $A$.
(c)* Does (a) remain true if $\mathbb{R}$ is replaced by $\mathbb{R}^n$?

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Ellipses from Circles

E 2914 [1981, 763]. *Proposed by R. C. Lyness, Southwold, England.*

A circle $B$ lies wholly in the interior of a circle $A$. $S$ is the set of all circles each of which touches $B$ externally and $A$ internally.

(i) Find the locus of the internal center of similitude of the pairs of circles from $S$. (ii) Prove that every point of the locus, except one, is the i.c.s. of exactly one pair of circles from $S$.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.*

(i) Let $R$ and $r$ stand for the radii of $A$ (center $M$) and $B$ (center $N$), respectively, and let $\sigma_1$ and $\sigma_2$ be members of the family $S$. The center and the radius of $\sigma_i$ will be denoted by $K_i$ and $\rho_i$ and the point of contact of $\sigma_i$ with $A$ and $B$ by $C_i$ and $D_i$ respectively ($i = 1,2$). Then we have:

$$\overline{MK}_i = R - \rho_i,\ \overline{NK}_i = r + \rho_i.$$

Hence $\overline{K_iM} + \overline{K_iN} = R + r$. The locus of the centers of the circles of $S$ is therefore the ellipse $e$ with $M$ and $N$ as its real foci and $R + r$ as the length of its major axis. The internal center of similitude $I$ of $\sigma_1$ and $\sigma_2$ lies therefore in the interior of $e$ and it may be conjectured that the locus of the internal centers of similitude of all pairs of $S$ is in fact the whole interior of $e$. To prove this we observe that $C_i$ is the external and $D_i$ the internal point of similitude of $\sigma_i$ with $A$ and $B$, respectively. Therefore $C_1$ and $C_2$ as well as $D_1$ and $D_2$ are collinear with the external point of similitude $E$ of $\sigma_1$ and $\sigma_2$. It is well known that $\sigma_1$ and $\sigma_2$ are interchanged under an inversion with center $E$ and a real circle of inversion. If $m$ is the radius of this circle, we have $\overline{EC}_1 \times \overline{EC}_2 = \overline{ED}_1 \times \overline{ED}_2 = m^2$. This means that $E$ has equal powers with respect to $A$ and $B$. The point $E$ is therefore on the radical axis $l$ of $A$ and $B$. Since $I$ and $E$ are harmonically separated by $K_1$ and $K_2$, we infer that $E$ is on the polar line $p$ of $I$ with respect to $e$.

(ii) Starting from any point $I$ in the interior of $e$, we can construct its polar line $p$ with respect to $e$. Its intersection $E$ with $l$ gives us the external point of similitude of *two* circles of $S$ having their centers on $IE$ (note that $E$ may be at infinity), *provided that $p$ does not coincide with $l$, that is: provided that $I$ does not coincide with the pole $L$ of $l$ with respect to $e$.* $L$ is the exceptional point mentioned in the problem. Any line through $L$ intersects $e$ in two points which are centers of circles belonging to $S$.

Also solved by J. Dou (Spain), L. Kuipers (Switzerland), W. A. Newcomb, A. Nijenhuis, and the proposer.

### A Triangle Inequality That Didn't Make It

E 2924* [1982, 63]. *Proposed by Jack Garfunkel, Flushing, NY.*

Triangle $A_1A_2A_3$ is inscribed in a circle; the medians through $A_1[A_2]$ meet the circle again at $M_1[M_2]$. The angle bisectors through $A_1[A_2]$ meet the circle again at $T_1[T_2]$. Prove or disprove: $|A_1M_1 - A_2M_2| \leqslant |A_1T_1 - A_2T_2|$.

*Solution by O. G. Ruehr, Michigan Technological University.* The conjecture is false, as is shown by the following counterexample. Take the circle to be the unit circle centered at the origin and let the triangle be the inscribed $30° - 60°$ right triangle with vertices

$$A_1 = (-1,0),\quad A_2 = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right),\quad A_3 = (1,0).$$

We find routinely that

$$T_1 = \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right),\quad T_2 = (0,-1),\quad \text{and } M_2 = \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right).$$

Then $A_1T_1 = A_2T_2 = \sqrt{2+\sqrt{3}}$ and $A_2M_2 = 2$. Clearly the segment $A_1M_1$ has length $< 2$ since it is not a diameter. Thus

$$|A_1M_1 - A_2M_2| > 0 = |A_1T_1 - A_2T_2|.$$

Also solved by L. Bankoff, W. A. Gilbert, H. J. Ludwig, and J. W. Reed.

**Geometric Closure and the Convex Hull**

E 2970 [1982, 698]. *Proposed by J. H. Conway, University of Cambridge, and B. Reznick, University of Illinois.*

Find all finite sets of points in the plane ($P = \{p_1, \ldots, p_n\}$) with the following "closure" property: If the line segments $\overline{p_i p_j}$ and $\overline{p_k p_l}$ intersect in a single point $p$, then $p \in P$.

*Solution by B. Richter (student), University of Waterloo, and H. Shank, Waterloo, Ontario, Canada.* It is easiest to characterize a finite set $S = \{p_1, \ldots, p_n\}$ having the closure property in terms of the number $m$ of vertices of its convex hull, $c(S)$.

CHARACTERIZATION. We must have $m \leqslant 4$ and

(1) if $m = 1$, then $n = 1$;

(2) if $m = 2$, then the points $p_i$ are all collinear;

(3) if $m = 3$, then either (i) $n = 6$ and we have the situation illustrated in Fig. 1; or (ii) there is a line $L$ through a vertex of $c(S)$ such that all the non-vertices of $S$ lie on $L$; or

(4) if $m = 4$, then $n \geqslant 5$; a fifth point is the intersection of the diagonals of $c(S)$ and all other points of $S$ lie on the same diagonal.

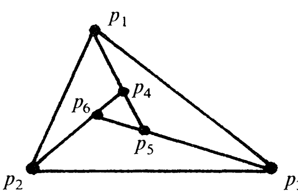



FIG. 1

The proof of this characterization revolves around the following two observations. The first is immediate.

INHERITANCE LAW. Let $S$ be a subset of the plane with the closure property and let $T$ be a subset of $S$. Then $S \cap c(T)$ has the closure property.

THEOREM. *Let $S$ be a finite subset of the plane. Suppose $S$ has the closure property and suppose, further, that $c(S)$ has four vertices. Let $p \in S$ be on the border of $c(S)$. Then $p$ is a vertex of $c(S)$.*

*Proof.* Let $v_1$, $v_2$, $v_3$ and $v_4$ be the four vertices of $c(S)$. Let them be labelled so that $p$ is on the side $v_1v_2$. Let $q$ be the intersection of the diagonals. (See Fig. 2.) Let $Q_0$ be $c(S)$. The segment $v_4p$ meets $v_1q$ in some point, say $q_1$, of $S$. Note that $q_1$ is in the interior of $Q_0$. Now consider the quadrilateral $Q_1$ determined by the points $p$, $v_2$, $v_3$ and $q_1$. The point $q$ lies on the segment $q_1v_3$. Thus we can repeat the argument to find a point $q_2$ in $S$ that is also in the interior of $Q_1$. Inductively, we can find quadrilaterals $Q_n$ and points $q_n$ in $S$ such that $q_n$ is a vertex of $Q_n$ and $q_{n+1}$ is in the interior of $Q_n$. But this implies that $S$ is not finite. Q.E.D.

COROLLARY A. *If $m = 4$, then any member of $S$ that is not a vertex lies on a diagonal.*

*Proof.* (See Fig. 3.) The quadrilateral $v_1pv_3v_4$ and the inheritance law would violate the theorem. Q.E.D.

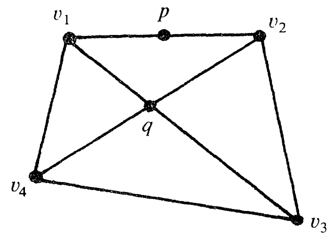



FIG. 2

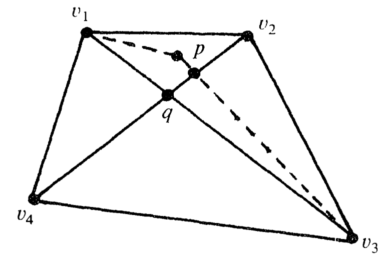



FIG. 3

COROLLARY B. *If* $m = 4$, *then any two members of* $S$ *that are not vertices lie on a common diagonal.*

*Proof*. (See Fig. 4.) Consider the quadrilateral $p_1p_2v_2v_1$. The diagonals meet at, say, $q_1$. Now $q_1q$ meets $p_1p_2$ at, say, $p_3$. Now $p_1p_2v_2v_1$ violates the theorem. Q.E.D.

Corollary B is (4) of the characterization.

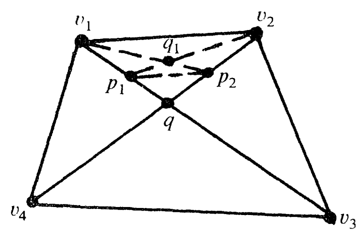



FIG. 4

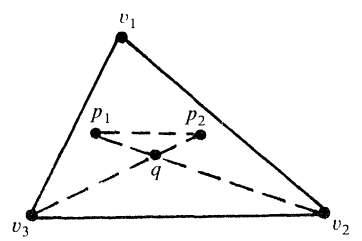



FIG. 5

COROLLARY C. *If* $m = 3$, *then any two points of* $S$ *that are not vertices of* $c(S)$ *lie on a line through a vertex.*

*Proof*. (See Fig. 5.) Let $q$ be the intersection of the diagonals $p_1v_2$ and $p_2v_3$. Let $q_1$ be the intersection of $v_1q$ and $p_1p_2$. Then $p_1p_2v_2v_3$ violates the theorem. Q.E.D.

From Corollary C, one can easily deduce (3).

COROLLARY D. *If* $S$ *is a finite set with the closure property, then* $c(S)$ *has at most four vertices.*

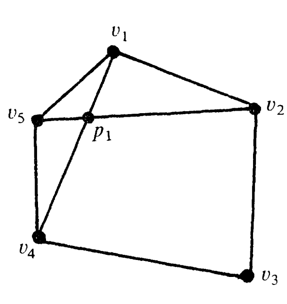



FIG. 6

*Proof*. (See Fig. 6.) Let $v_1, \ldots, v_5$ be vertices of $c(S)$. Let $p_1$ be the intersection of $v_1v_4$ with $v_2v_5$. Observe that the quadrilateral $v_2v_3v_4v_5$ violates the theorem. Q.E.D.

The cases (1) and (2) of the characterization are trivial. It is also straightforward to verify that

any set of points satisfying any one of (1) through (4) is indeed a finite subset of the plane, with the closure property.

Also solved by W. Janous (Austria), O. P. Lossers (The Netherlands), J. G. Mauldon, V. Pambuccian (Romania), R. Patenaude, P. J. Ryan (Canada), A. J. Schwenk, E. Uerhoeff (The Netherlands), and the proposer.

### An Essentially Circular ODE

E 2971 [1982, 698]. *Proposed by Barry L. Zaslove, Northeastern University.*

Find the solution curves of the ODE,

$$y''' = y'(3y''^2 - y''' y').$$

*Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* The differential equation may be written as:

$$\frac{y'''(1+y'^2) - 3y'y''^2}{(1+y'^2)^{5/2}} = 0,$$

where the left-hand member is obviously the derivative of

$$\frac{y''}{(1+y'^2)^{3/2}},$$

i.e., of the curvature $K$ of the solution curves. Hence $K$ is constant and the solution curves are the straight lines and the circles of the plane, with the exception of the lines $x$ = constant.

Also solved by 51 other readers and the proposer.

### A Problem Due to Steiner

E 2992 [1983, 286]. *Proposed by Jordi Dou, after a suggestion of the late M. S. Knebelman.*

Find the shape of a contour of length $L$ that encloses the largest possible area and is constrained to pass through three given points.

*Comment by M. S. Klamkin, University of Alberta, Canada.* This problem is due to Steiner (see J. Steiner, Gesammelte Werke, II, pp. 75–91) and has also been treated in my Olympiad Corner (Crux Mathematicorum, 6 (1980) 280–282).

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1986. The solver's full post-office address should be on each sheet.*

6499. *Proposed by Doug Hensley, Texas A & M University.*

Suppose $X_1, X_2, \dots$ are independent random variables all uniformly distributed on $[0,1]$. Show that

$$\lim_{n\to\infty} \operatorname{Prob}\left(\sum_{k=1}^{n} e^{-nX_k} \leqslant 1\right) = e^{-\gamma},$$

where $\gamma$ is Euler's constant.

6500. *Proposed by Eugene M. Luks, University of Oregon, and Michael B. Ward, Bucknell University.*

In the first printing of an algebra text, the following problem appeared:

Let $G$ be a group, and $H$ a normal subgroup of $G$. Prove: If every element of $G/H$ has a square root, and every element of $H$ has a square root, then every element of $G$ has a square root.

The statement is true if $G$ is abelian or finite. Is it true if $G/H$ and $H$ are abelian? Is it true if $G$ is torsion? Is it true in general?

---

**Do you have a problem?** Submit it to the Advanced Problems Section. This section is running short of challenging and attractive problems. Advanced problems should have reasonably wide appeal and have non-tedious solutions. They may involve results unfamiliar to senior undergraduate students, but problems that are too technical or too specialized or whose solutions require arcane, little-known results are not in general suitable.

---

## SOLUTIONS OF ADVANCED PROBLEMS

### A Metric Space Having a Given Group of Isometries

6398 [1982, 503; 1984, 146]. *Proposed by Edgar Feldman and Michael Vulis, CUNY Graduate Center, New York.*

Is it true that for any group $G$ of cardinality less than or equal to that of the continuum, one can construct a metric space $M$ such that $G$ is its full group of isometries? *What if $|G|$ has greater power?

*Solution by Randall Dougherty (graduate student), University of California, Berkeley.* Such a metric space $M$ exists no matter how large $G$ is.

Given $G$, choose a one-to-one function $F$ mapping $G$ into some limit ordinal $\lambda$. Let $z$ be any object which is not in $G$. Choose $(G \cup \{z\}) \times \lambda$ as the underlying set for the metric space $M$. Define the distance function $d$ for $M$ as follows: if $\alpha \neq \beta$, let $d((z,\alpha),(z,\beta)) = 1.1$; if $a \in G$ and $\alpha \leqslant \beta$, let $d((a,\alpha),(z,\beta)) = 1.2$; if $a \in G$ and $\alpha > \beta$, let $d((a,\alpha),(z,\beta)) = 1.3$; if $a \in G$ and $\alpha \neq \beta$, let $d((a,\alpha),(a,\beta)) = 1.4$; if $a, b \in G$, $a \neq b$, and $F(a^{-1}b) = \alpha$, let $d((a,\alpha),(b,\alpha+1)) = 1.5$; if $d(p,q)$ has not yet been determined, let $d(p,q) = 1.6$.

First note that the set $\{(z,\alpha)\colon \alpha \in \lambda\}$ is distinguishable in $M$ (using the 1.1 distances), and hence fixed under all isometries of $M$. Now, a point in $M$ has the form $(a,0)$ with $a \in G$ if and only if it is at distance 1.2 from all points $(z,\alpha)$; the point $(z,0)$ is the only point in $M$ at distance 1.3 from all points in $M$ other than those of the form $(z,\alpha)$ or $(a,0)$; the points $(a,1)$, $a \in G$, are just those points at distance 1.3 from $(z,0)$ but at distance 1.2 from all other points $(z,\alpha)$; and so on. In general, for any fixed $\beta < \lambda$, the points $(a,\beta)$, $a \in G$, are just those points at distance 1.3 from all $(z,\alpha)$ with $\alpha < \beta$, but at distance 1.2 from all other points $(z,\alpha)$; and $(z,\beta)$ is the unique point not in $\{(z,\alpha)\colon \alpha < \beta\}$ which is at distance 1.3 from all points of the form $(a,\alpha)$ with $a \in G$ and $\alpha > \beta$. Therefore, by transfinite induction, for each $\beta < \lambda$, the set $\{(a,\beta)\colon a \in G\}$ and the point $(z,\beta)$ are fixed under all isometries of $M$.

Using the 1.4 distances, we now see that for any isometry $i$ of $M$, there is a permutation $\pi$ of $G$ such that $i((a,\alpha)) = (\pi a,\alpha)$ and $i((z,\alpha)) = (z,\alpha)$ for all $a \in G$ and $\alpha < \lambda$. Let $g = \pi e$, where $e$

is the identity element of $G$; then $\pi e = ge$. Given any $x \in G$ other than $e$, let $\alpha = F(x)$. Then $d((e,\alpha),(x,\alpha+1)) = 1.5$, so $d((\pi e,\alpha),(\pi x,\alpha+1)) = 1.5$, so $F((\pi e)^{-1}\pi x) = \alpha$; since $F$ is one-to-one, $x = (\pi e)^{-1}\pi x = g^{-1}\pi x$, so $\pi x = gx$. Conversely, given any $g \in G$, it is easy to see that the permutation $\pi$ defined by $\pi x = gx$ induces an isometry of $M$. Therefore, $M$ has $G$ as its full group of isometries.

Also solved by M. J. Pelling (England). A solution by the proposers to the first part of the problem (i.e., with $G$ having cardinality less than or equal to that of the continuum) was published in this MONTHLY, 91 (1984) 146, #6398.

### Two Series

6450 [1984, 59–60]. *Proposed by J. O. Shallit, University of California, Berkeley.*

(a) Show that

$$\sum_{k=1}^{\infty} \frac{k - n[k/n]}{k(k+1)} = \log n,$$

where $n$ is a positive integer.

(b) Let $S(n,k)$ denote the sum of the digits of $k$ when expressed in base $n$. Evaluate

$$\sum_{k=1}^{\infty} \frac{S(n,k)}{k(k+1)}.$$

*Composite solution.*

(a)
$$\begin{aligned}\sum_{k=1}^{nN-1} \frac{k-n[k/n]}{k(k+1)} &= \sum_{k=1}^{nN-1} \frac{1}{k+1} - n\sum_{j=1}^{N-1}\sum_{k=jn}^{(j+1)n-1} \frac{j}{k(k+1)} \\ &= \sum_{k=1}^{nN-1} \frac{1}{k+1} - n\sum_{j=1}^{N-1} j\left(\frac{1}{nj} - \frac{1}{(j+1)n}\right) \\ &= \sum_{k=1}^{nN-1} \frac{1}{k+1} - \sum_{j=1}^{N-1} \frac{1}{j+1} \to \log n \quad \text{as } N \to \infty.\end{aligned}$$

(b) If we write

$$k = \sum_{i=0}^{\infty} c_i n^i, \quad 0 \leqslant c_i < n,$$

then

$$[kn^{-j}] = \sum_{i=j}^{\infty} c_i n^{i-j},$$

so that

$$\begin{aligned}\sum_{j=1}^{\infty} [kn^{-j}] &= \sum_{i=1}^{\infty} c_i \sum_{j=1}^{i} n^{i-j} = \frac{1}{n-1}\sum_{i=1}^{\infty} c_i(n^i - 1) \\ &= \frac{1}{n-1}(k - S(n,k)).\end{aligned}$$

It follows that

$$S(n,k) = k - (n-1)\sum_{j=1}^{\infty} [kn^{-j}]$$

$$= (n-1)\sum_{j=1}^{\infty} n^{-j}(k - n^j[kn^{-j}]),$$

and consequently, by (a), that

$$\sum_{k=1}^{\infty}\frac{S(n,k)}{k(k+1)} = (n-1)\sum_{j=1}^{\infty} n^{-j}\sum_{k=1}^{\infty}\frac{k-n^j[kn^{-j}]}{k(k+1)}$$
$$= (n-1)\sum_{j=1}^{\infty} n^{-j}\log n^j$$
$$= \frac{n}{n-1}\log n.$$

Solved by the proposer and forty-four others.

**A Function Satisfying Certain Conditions**

6451 [1984, 144]. *Proposed by C.L. Mallows, Bell Laboratories, Murray Hill, NJ.*

Construct a function $f$ on $[0,\infty)$ that satisfies $f(0)=0$, $f(1)=1$, $f(x)=f(2-x)$ for $0<x<2$, $f'(x)=2f(2x)$ for $x>0$, and $f(x)\not\to 0$ as $x\to\infty$.

*Solution by Thøger Bang, University of Copenhagen, Copenhagen, Denmark.* We shall prove that there exists a unique continuous function $f$ on $(0,\infty)$ satisfying $f(1)=1$, $f(x)=f(2-x)$ for $0<x<2$, and $f'(x)=2f(2x)$ for $x>0$. The other conditions are redundant, for if $f$ satisfies the above conditions, then by induction,

$$f(x) = (-1)^n f(2^{n+1}-x) \quad \text{for } n = 0,1,2,\dots, 0 < x < 2^{n+1},$$

and consequently the range of $|f(x)|$ on $[m, m+1]$ is independent of the integer $m$; it follows that $f(x)\not\to 0$ as $x\to\infty$ and, by continuity, that $f(0)=f(2)=0$. In addition, such a function satisfies

$$f(x) = \int_0^{2x} f(t)\,dt \quad \text{for } x > 0$$

and, since

$$2f(2x) = 2f(2-2x) = f'(x) = f'(1-x),$$

it also satisfies $f(x)+f(1-x)=1$ for $0\leqslant x\leqslant 1$.

Now consider a function $g$ such that

(1) $$g \text{ is continuous on } [0,1] \quad \text{and } g(0)=0,$$

(2) $$g(x)+g(1-x) = 1 \quad \text{for } 0\leqslant x\leqslant 1.$$

This function can be uniquely extended to a function on $[0,\infty)$ satisfying

(3) $$g(x) = (-1)^n g(2^{n+1}-x) \quad \text{for } 0\leqslant x\leqslant 2^{n+1}, \quad n=0,1,2,\dots.$$

We then also have $\int_0^{2^{n+2}} g(t)\,dt = 0$ for $n=0,1,2,\dots$.

Denote by $S$ the space of functions $g$ on $[0,\infty)$ satisfying (1), (2) and (3). Then $S$ is a closed subset of the Banach space of bounded continuous functions on $[0,\infty)$ with the uniform norm. Given any $g\in S$, define

$$g^*(x) = \int_0^{2x} g(t)\,dt \quad \text{for } x\geqslant 0.$$

Then $g^*(0)=0$ and, for $0\leqslant x\leqslant 1$,

$$g^*(x) + g^*(1-x) = 2\int_0^1 g(t)\,dt = 1.$$

Also, for $0 \leqslant x \leqslant 2^{n+1}$, $n = 0,1,2,\ldots,$

$$g^*(x) = \int_0^{2x} g(t)\,dt = (-1)^n \int_0^{2^{n+2}-2x} g(t)\,dt = (-1)^n g^*(2^{n+1}-x). \tag{4}$$

Hence $g^* \in S$. Moreover, since $g^*(0) = 0 = g(0)$ and $g^*(1/2) = 1/2 = g(1/2)$, it follows from (4) that, for $n = 0,1,2,\ldots, g^*(n/2) = g(n/2) = c_n$, where $c_n$ is independent of $g$.

Suppose now that $g_1, g_2 \in S$. Then, for $x \geqslant 0$,

$$g_1^*(x) - g_2^*(x) = \int_n^{2x} (g_1(t) - g_2(t))\,dt,$$

where $n$ is the integer nearest to $2x$. Hence, for the uniform norm,

$$\|g_1^* - g_2^*\| \leqslant \frac{1}{2}\|g_1 - g_2\|$$

and so the *-operation is a contraction. It follows that there is a unique $f \in S$ such that $f^* = f$, and this $f$ has the required properties.

Also solved by Patrick J. Fitzsimmons, Tomas P. Schonbek, V. Daniel Zurkowski, and the proposer.

### A Sequence of Random Variables

6452 [1984, 144]. *Proposed by Giovanni Emmanuele and Alfonso Villani, Catania, Italy.*

Construct a sequence of nonnegative random variables $\{\xi_n\}_{n\geqslant 1}$ such that $\sup_{n\geqslant 1}\mathbb{E}\xi_n^p < \infty$ for all $p > 0$, and $\mathbb{P}(\sup_{j\geqslant 1}\xi_{n_j} < \infty) = 0$ whenever

$$1 \leqslant n_1 < n_2 < \cdots < n_j < \ldots.$$

*Solution by Lajos Takács, Case Western Reserve University, Cleveland, Ohio.* Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be a sequence of mutually independent random variables for which

$$\mathbb{P}\{\xi_n \leqslant x\} = 1 - e^{-x} \quad \text{if } x \geqslant 0.$$

Then

$$\mathbb{E}\{\xi_n^p\} = \int_0^\infty x^p e^{-x}\,dx = \Gamma(p+1) < \infty$$

for all $p > 0$. In this case

$$\mathbb{P}\Big\{\max_{1\leqslant j\leqslant k} \xi_{n_j} \leqslant x\Big\} = (1 - e^{-x})^k$$

for $x \geqslant 0$ and $k = 1,2,\ldots$. If we let $k \to \infty$, then by the continuity theorem for probabilities we obtain that

$$\mathbb{P}\Big\{\sup_{j\geqslant 1}\xi_{n_j} \leqslant x\Big\} = 0 \quad \text{for all } x \geqslant 0.$$

If we let $x \to \infty$, then again by the continuity theorem for probabilities we obtain that

$$\mathbb{P}\Big\{\sup_{j\geqslant 1}\xi_{n_j} < \infty\Big\} = 0.$$

Also solved by A. N. Al-Hussaini & J. R. McGregor (Canada), L. E. Clarke (England), Alejandro deAcosta, Victor Hernández (Spain), Ellen Hertz, Ignacy Icchak Kotlarski, O. P. Lossers (Netherlands) and the proposers.

### An Exponential Identity

6453 [1984, 144]. *Proposed by Richard Askey, University of Wisconsin, Madison.*

Show that, for any real $x$,

$$\lim_{\lambda\to\infty}\prod_{n=0}^{\infty}\left(1+\frac{\lambda x^2}{(n+\lambda)^2}\right)=e^{x^2}.$$

*Composite solution.* For $\alpha>0$, $\lambda>0$ we have

$$\int_0^\infty \frac{dt}{(t+\lambda)^{1+\alpha}} \leqslant \sum_{n=0}^{\infty}\frac{1}{(n+\lambda)^{1+\alpha}} \leqslant \frac{1}{\lambda^{1+\alpha}}+\int_0^\infty \frac{dt}{(t+\lambda)^{1+\alpha}}$$

and hence

$$\lim_{\lambda\to\infty}\sum_{n=0}^{\infty}\frac{\alpha\lambda^{\alpha}}{(n+\lambda)^{1+\alpha}}=1.$$

It follows that if $z$ is any complex number and $\varepsilon>0$, then

$$\sum_{n=0}^{\infty}\log\left(1+\frac{\varepsilon\lambda^{\varepsilon}z}{(n+\lambda)^{1+\varepsilon}}\right)=\sum_{n=0}^{\infty}\frac{\varepsilon\lambda^{\varepsilon}z}{(n+\lambda)^{1+\varepsilon}}+O\left(\sum_{n=0}^{\infty}\frac{\varepsilon^2\lambda^{2\varepsilon}|z|^2}{(n+\lambda)^{2+2\varepsilon}}\right)$$
$$\to z \quad \text{as } \lambda\to\infty$$

and consequently

$$\lim_{\lambda\to\infty}\prod_{n=0}^{\infty}\left(1+\frac{\varepsilon\lambda^{\varepsilon}z}{(n+\lambda)^{1+\varepsilon}}\right)=e^{z}.$$

This is more general than the identity in the problem.

Solved by the proposer and fifty others.

### Evaluation of an Integral

6454 [1984, 205]. *Proposed by the Chico Problem Group, California State University, Chico.*

Evaluate the integral

$$I(p)=\int_0^{\pi/2}\theta\tan^{p}(\theta)\,d\theta,\quad -2<p<1,$$

in terms of the digamma function.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, Netherlands.* The substitution $\tan\theta=x$ yields

$$I(p)=\int_0^\infty\frac{x^p}{1+x^2}\tan^{-1}x\,dx.$$

Let

$$J(a)=\int_0^\infty\frac{x^p}{1+x^2}\tan^{-1}(ax)\,dx.$$

Then, for $-2<p<0$,

$$J'(a)=\int_0^\infty\frac{x^{p+1}\,dx}{(1+x^2)(1+a^2x^2)}=\frac{1}{2(1-a^2)}\left(\int_0^\infty\frac{x^{p+1}}{1+x^2}\,dx-\int_0^\infty\frac{a^2x^{p+1}}{1+a^2x^2}\,dx\right).$$

Substituting $x = \sqrt{t}$ and $x = \sqrt{t}/a$, respectively, in the last two integrals yields, for $-2 < p < 0$,

$$J'(a) = \frac{1 - a^{-p}}{2(1 - a^2)} \int_0^\infty \frac{t^{p/2}}{1 + t}\,dt = \frac{\pi}{2\sin(p\pi/2)} \cdot \frac{a^{-p} - 1}{1 - a^2}$$

(see formula 3.241.2 in [**1**]). Hence, for $-2 < p < 0$,

$$\begin{aligned} I(p) = J(1) &= \frac{\pi}{2\sin(\pi p/2)} \int_0^1 \frac{t^{-p} - 1}{1 - t^2}\,dt \\ &= \frac{\pi}{4\sin(\pi p/2)} \int_0^1 \frac{u^{-(p+1)/2} - u^{-1/2}}{1 - u}\,du \\ &= \frac{\pi}{4\sin(p\pi/2)}\left(\psi\left(\frac{1}{2}\right) - \psi\left(\frac{1-p}{2}\right)\right), \end{aligned}$$

where $\psi$ is the digamma function (see formula 3.231.5 in [**1**]). By analytic continuation the identity in fact holds for $-2 < p < 1$ with

$$I(0) = \lim_{p \to 0} I(p) = \pi^2/8.$$

**Reference**

1. I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series and Products, Academic Press, 1965.

Also solved by Paul F. Byrd, M. L. Glasser, A. A. Jagers (Netherlands), Frederic R. Schwab, and the proposers.

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Linear Algebra and Group Representations, Volumes 1 and 2*. By Ronald Shaw. Academic Press, New York, 1982. Vol. 1, xi + 269 pp., \$32.00. Vol. 2, ix + 270–579 pp., \$35.00.

GERALD J. JANUSZ
*Department of Mathematics, University of Illinois, Urbana, IL 61801*

The representation theory of groups is studied because an abstract group, initially defined by a set of axioms, takes on a concrete existence when its elements are described as linear transformations on a vector space over some field. The structure implied by the group axioms is intertwined with the structure of the vector space and all its related objects to give insight into properties of groups of linear transformations and of abstract groups in general. What could be more natural than to study the subject of linear (and multilinear) algebra simultaneously with the subject of representations of abstract groups as groups of linear transformations? In this book (I regard the two volumes as one book) the rich subject of linear algebra is laid before the reader not only as a subject worthy of study in its own right but also with the goal of providing a framework in which the basic results about representations of abstract groups can be efficiently proved. The abstract groups involved are not always assumed to be finite, but the emphasis is placed upon algebraic results rather than topological, analytic, or arithmetic results.

What is the structure possessed by a vector space and what are the related objects? The beginning student might be overwhelmed by Professor Shaw's reply: dual spaces, covariant and

contravariant tensors, the tensor, symmetric, and exterior algebras, dyads, projections, the various types of bilinear forms, hermitian forms, isometries, etc., etc. However, the determined and questioning student will be rewarded by a study of linear algebra firstly because the ideas of this subject are in the toolbox of every theoretical scientist, and secondly, because of the intrinsic elegance of the subject. Shaw gives a highly polished treatment using some ideas not commonly seen in other texts. The systematic use of dyads (rank one linear transformations) for example, permits convenient descriptions of projection maps and decompositions of the identity map. The elegance of this treatment will be apparent to the reader who already knows the subject but, unfortunately, may be lost upon (and even confusing to) the beginning student. The presentation assumes considerable sophistication on the part of the reader. In somewhat less than 350 pages of material on linear algebra, Shaw covers far more material than other texts of comparable length. This is achieved by leaving many important results to the exercises (with ample hints). In return for this price, the reader is treated to examples and illustrations drawn from Shaw's background in mathematical physics.

Consider, for example, the Minkowski space; this is a four-dimensional vector space $M$ equipped with a bilinear form with signature $(1,3)$. The associated quadratic form $q = x^2 - y^2 - z^2 - w^2$ is not definite and so one seldom finds a detailed discussion of it in standard texts. In this text it is treated in great detail, as is its group of isometries, the Lorentz group, because of the connections with relativity and quantum mechanics.

One of the tools used throughout the text is the tensor algebra and the closely related symmetric and exterior algebras. For a vector space $V$ over a field $F$, the tensor algebra $T(V)$ is the graded $F$-algebra which has the $r$-fold tensor product of $V$ with itself as the $r$th homogeneous component. From it one obtains the symmetric algebra as a homomorphic image (essentially a polynomial ring) and the exterior algebra, a finite dimensional algebra having a distinguished one-dimensional subspace—an ideal even. These algebras are functorial in $V$; thus a linear transformation on $V$ induces an automorphism on each of the algebras and moreover, the grading of each is preserved. This gives, for example, an action of $\mathrm{GL}(V)$ upon the distinguished one-dimensional ideal of the exterior algebra by which one defines the determinant map. One sees universal objects and commutative diagrams—a very modern approach to this very old subject.

The remaining 200 pages or so are devoted to various results about representations of finite and infinite groups. There is an interesting mixture of subject matter and of style in these pages. One finds many standard and basic results about representations of finite groups often presented in a way that does not use finiteness. But there is much more! Here are two examples of not so standard topics.

Let $D$ denote the natural representation of the special linear group, $G = \mathrm{SL}(V)$, of a two-dimensional complex vector space $V$. If $\{x, y\}$ is a basis for $V$, then the $n$th symmetric power $S_n$, of $V$ may be viewed as the space of homogeneous polynomials of degree $n$ in the variables $x$ and $y$. The representation of $G$ on $S_n$ is denoted by $D^n$. (This notation differs slightly from Shaw's.) We have thereby constructed an infinite list of irreducible representations of $G$. Now consider the tensor product, $D^m \otimes D^n$, of two of these representations. How does it decompose? The solution to this problem is elegant. As a special case of a more general result, one knows that the vector space of $G$-homomorphisms from $S_k$ into $S_m \otimes S_n$ has the same dimension as the space of $G$-invariant elements in $S_k \otimes S_m \otimes S_n$. Next one finds such a nonzero invariant element for fixed $m$, $n$ and all $k$ on a certain interval. This shows that the corresponding $D^k$ are constituents of the tensor product $D^m \otimes D^n$. By an act of God, these $D^k$ have different degrees which sum to the degree of $D^m \otimes D^n$ and so the decomposition is accomplished. The details are highly nontrivial and make heavy use of algebraic computations with dual bases. The study of these invariant elements goes back to Clebsch in 1872 and is in sharp contrast to the modern ideas mentioned above.

Another pretty result is the isomorphism between the orthogonal group of positive definite three-dimensional form over the complex field and the quotient of the two-dimensional group of isometries of an alternating form modulo its center, the projective symplectic group. This

well-known isomorphism is efficiently proved by making use of the analysis of the Minkowski space mentioned above.

One of the most attractive features of complex representation theory of finite groups is the theory of characters along with the arithmetic results which can be obtained by exploiting the fact that the values of the characters on group elements are algebraic integers. One such result is the fact that the degrees of the irreducible characters divide the group order. It is unfortunate that Shaw does not give more than token attention to characters and does not mention any arithmetic results such as this one. In fairness it must be said that Shaw's goals do not suffer seriously from these omissions.

I am very pleased with this book, especially with the linear algebra portions. It provides an alternative approach to many basic results in the subject where new insights are not easily acquired. The applications motivated by the interests of a mathematical physicist provide new insights to one educated in the traditional approach. The author states that the book is intended for advanced students. I would go a bit farther and state that the book is excellent for the student who is already quite knowledgeable in the subject and desires additional insights and applications.

---

*Riemannian Geometry*. By Wilhelm Klingenberg. Walter de Gruyter, Berlin, Germany, 1982. x + 396 pp.

H. WU
*Department of Mathematics, University of California, Berkeley, CA 94720*

Writing a readable introduction to differential geometry is among the most difficult feats in the art of mathematical exposition. Such an introduction should ideally fulfill the dual purpose of preparing beginners for *most* contingencies in their march to the frontier of research, and at the same time *educating* them in the ways of geometric life without ramming facts down their throats.

Now, since any mathematical work exceeding 1000 pages (perhaps even less) automatically invites suspicion, there would also be a natural upper bound on its size. This being the case, the two preceding boundary conditions may well be inherently incompatible. At the very least, no one has yet succeeded in simultaneously achieving these two objectives and, given the uncontrolled growth of the field in the past decade, the chance of anyone succeeding in the future would seem even more remote.

The eminently quotable Michael Spivak, as one who has thought deeply about such things, explains the predicament facing any author in geometry:

> "For many years I have wanted to write the Great American Differential Geometry book. Today a dilemma confronts anyone intent on penetrating the mysteries of differential geometry. On the one hand, one can consult numerous classical treatments of the subject in an attempt to form some idea how the concepts within it developed. Unfortunately, a modern mathematical education tends to make classical mathematical works inaccessible. On the other hand, one can now find texts as modern in spirit, and as clean in exposition, as Bourbaki's algebra. But a thorough study of these books usually leaves one unprepared to consult classical works, and entirely ignorant of the relationship between elegant modern constructions and their classical counterparts. Most students eventually find that this ignorance of the roots of the subject has its price–no one denies that modern definitions are clear, elegant, and precise; it is just that it's impossible to comprehend how anyone ever thought of them. And even after one does master a modern treatment of differential geometry, other modern treatments often appear to be about totally different subjects." ([**12**], Volume I, p.i.)

On this pessimistic note then, let us turn to a brief examination of some of the existing works of this genre in order to put the volume under review in perspective.

The Spivak opus [**12**] is a five-volume work totalling 2643 typed pages (which would come to about 1000 printed pages). Their great length notwithstanding, these books most emphatically are introductory texts. As an example, Spivak's attitude towards curvature is that "... there is no point in introducing the curvature tensor without explaining how it was invented and what it has to do with curvature" (p.i. in Volume I of [**12**]). The background he assumed of the reader is minimal: advanced calculus, linear algebra and metric spaces. But because of this he had to pay a price: out of the 635 typed pages of Volume I, 546 are devoted to such preparatory material as differentiable manifolds, tensor fields, Lie groups, algebraic topology.... I point this out not as a criticism, far from it, for I know of no other author willing to perform this valuable service for the beginner *and* who still goes on to write about substantive geometry. In addition, and I believe this is Spivak's great contribution to book-writing, the entire Volume II (403 typed pages) is devoted to discussing not only the contributions of Gauss and Riemann, but also how they tie in with such latter-day constructs as moving frames and principal fibre bundles. This material is available nowhere else. The remaining three volumes (1605 typed pages) are concerned with more advanced topics which culminate in characteristic classes and the Gauss-Bonnet theorem in $n$-dimensions. There is the same clarity and insight all the way to the end. So with all these going for it, why then is Spivak [**12**] not *THE* book in differential geometry? The reason is that Spivak, being ever so solicitous of the reader's learning process, would not allow himself to cram the maximal amount of material into a fixed number of pages purely for the sake of cramming. Thus, long as it is, a few topics of vital importance have been left out of [**12**], and among these are: homogeneous and symmetric spaces, and Hermitian and Kähler manifolds.

By contrast, these and more can be found in the two-volume work of Kobayashi-Nomizu [**8**]. This is the work that convincingly demonstrates what efficiency and expertise can accomplish in 777 printed pages. Minor carping from some quarters aside, this is *THE* standard reference work of the subject. By the same token, anyone who thrusts these two volumes on a beginner for use as an introductory text would be guilty of committing an act of inhumanity against a fellow being. In the terse 62-page Chapter 1, for example, the prerequisites for reading this work are relentlessly itemized (with some proofs): differentiable manifolds, tensor fields, Lie groups and fibre bundles. Of course the official line is that anyone not acquainted with this material can pick it up from there, but let no one be so deceived! Moreover, given the reference nature of the work, there is no room in this kind of tightly organized writing for such activities as giving motivation or tracing the historical evolution of a concept. (After all, does anyone expect the Oxford English Dictionary to be a treatise on etymology?) But when all is said and done, this work is the one place where one can find a logical and systematic development of differential geometry *ab initio* (à la Bourbaki) which does lead to the frontier of research (at the time of its writing) in a great many areas. For that alone we have to be thankful.

At this point, I hope the reader may be more receptive to the idea of seeking instead a readable, compact introduction to a *specific area* in differential geometry. Since Riemannian geometry is at the core of differential geometry no matter how the latter is defined, it comes as no surprise that indeed several attempts have been made to write textbooks on precisely this topic. For the purpose of this review, it suffices to narrow our scope even further. We will be discussing *synthetic Riemannian geometry*, which may be loosely defined as the area where theorems are proved by predominantly geometric arguments, with "bare hands" so to speak, and where hard analysis intrudes but slightly. A typical theorem here is the sphere theorem (Rauch, Berger, Klingenberg), which states that a compact simply-connected Riemannian manifold whose sectional curvature $K$ satisfies $1/4 < K \leqslant 1$ is homeomorphic to the sphere. The proof involves nothing more than calculus and point-set topology, plus a generous supply of ingenuity and geometric insight.

Now it so happens that there are already two classic expositions of synthetic Riemannian

geometry: Gromoll-Klingenberg-Meyer [**3**] and Cheeger-Ebin [**1**]. I shall refer to these as GKM and CE respectively. Starting with minimal prerequisites (for GKM, the same as Spivak; for CE, the first half of GKM, say), they lead the reader straight to the deepest available theorems in synthetic Riemannian geometry at the time of their respective publications: for GKM, this would mean the sphere theorem, and for CE, the differentiable version of the sphere theorem plus the structure theorems of noncompact manifolds of nonnegative curvature and compact manifolds of nonpositive curvature. By design, neither book deals with such topics as connections on fibre bundles, differential forms and characteristic classes, or Kähler metrics. In GKM, historical references are sprinkled throughout the text but are nowhere pursued as vigorously as in Spivak; CE is even more reticent in such matters. However, the expository style of both books possesses that rare combination of clarity, incisiveness, and power. Within the confines of their avowed goals, both are unqualified successes.

I may add that the simple beautiful German of GKM is much easier to read than quite a few English mathematical texts that have come my way. Thus the linguistic problem is hardly a serious one.

And finally, we come to the volume under review (all 396 densely-printed pages of it). It represents the author's second thoughts on GKM, so to speak. What he succeeded in doing is to subsume all of GKM as a proper subset of the first two chapters of this volume (exactly two-thirds in length), and to devote the third and final chapter to a systematic account of the structure of the geodesic flow. The prerequisites are the same as Spivak or GKM (with occasional exceptions). Unlike any of the above-mentioned works, this book begins with the definition of manifolds modelled on Hilbert space. This makes possible a lengthy discourse on the Hilbert manifold of curves, a topic that is appearing in an elementary text on Riemannian geometry for the first time.

The style of this book is noticeably different from that of GKM: more chatty, the notation at times more formidable, and, surprisingly, in some places overtaxing the reader's perseverance and good will. But whatever the loss vis-à-vis the virtues of GKM (inevitable in any such expansion, be it a restaurant or a book), there is ample compensation in being able to include some delectable topics from the theory of geodesic flows. Here one finds a proof of the theorem of Lyusternik-Schnirelman (a Riemannian manifold homeomorphic to the 2-sphere must possess three simple closed geodesics), an account of Anosov's work on manifolds of negative curvature, and a long detailed discussion of closed geodesics on compact surfaces in general and on ellipsoids in particular. There is also a proof of a beautiful theorem of E. Hopf: *Any metric without conjugate points on the* 2*-torus must be flat*. A generalization of this to higher dimensions would be most interesting.

In a general review of this nature, any technical discussion would be out of place. Yet I consider the omission of the *definitions* of the Ricci and scalar curvatures in such an introductory text to be serious enough to warrant a breach of protocol on my part. This is serious because, not only are these curvatures among the most basic objects in Riemannian geometry, but also very great progress has just been made in our understanding of (specifically) these two curvatures in the past few years (see, for example, [**4**]–[**6**], [**9**]–[**11**] and [**13**]). There is no excuse for not letting the reader in on something of such great current interest. The omission of Ricci curvature hampers the author's exposition as well: the Bonnet-Myers theorem (which is incidentally left out of the Index but which appears on p. 205) is now stated in terms of the sectional curvature rather than the Ricci curvature. One objects to this not because of the obvious loss of generality but rather on conceptual grounds. Perhaps this omission also accounts for the author's neglect to mention on p. 229 that the Maximal Diameter Rigidity Theorem of Toponogov has been generalized by S. Y. Cheng [**2**] to a statement concerning Ricci curvature; elsewhere, the author is very conscientious about pointing out the latest advances concerning some of the theorems in the book. (On this same subject, mention should be made of some curious omissions and errors in the references to the literature on pp. 239–40 and on p. 254.)

To return to the first topic of this review, it is unlikely, as of 1984, to see a textbook introducing the student to all or even most of the topics of current interest in differential geometry. The field has simply grown too big (see the introduction to Volume III of [**12**], which was written back in 1975!). Even if one starts where [**8**] leaves off, would it be reasonable to expect an author to write about minimal submanifolds, harmonic maps, spectra of compact manifolds, Yang-Mills fields, function theory on non-compact manifolds, holomorphic mappings, holomorphic vector bundles over complex projective spaces, applications of the Monge-Ampère equations, etc., all within the guidelines spelled out at the beginning of this review? The best one can hope for is perhaps a systematic and concise account of a given topic. In this sense, the present volume, together with the author's other book [7], fill a gap in the area of geodesic flows and closed geodesics. In the author's words: "I have no other excuse to proffer for the selection of contents, except that I am convinced that my choice represents a lasting contribution to the field, and that future fruitful developments seem most likely" (p. vii). What is arguably the most urgent need at present is an introduction to the recent advances in harmonic maps, minimal submanifolds, Yang-Mills fields and the geometric applications of the Monge-Ampère equation; these seemingly disparate topics are tied together by one common thread, that of nonlinear elliptic equations, which is itself in need of a good introductory text. Is there no one to take up the challenge?

**References**

**1.** J. Cheeger and D. Ebin, Comparison Theorems in Riemannian Geometry, North Holland Publishing Co., Amsterdam and Oxford, 1975.

**2.** S. Y. Cheng, Eigenvalue comparison theorem and its geometric applications, Math. Z., 143 (1975) 289–297.

**3.** D. Gromoll, W. Klingenberg, and W. Meyer, Riemannsche Geometrie im Grossen, Lecture Notes in Mathematics 55, Springer-Verlag, Berlin-Heidelberg-New York, 1968.

**4.** M. Gromov and H. B. Lawson, Spin and scalar curvature in the presence of a fundamental group, Ann. Math., 111 (1980) 209–230.

**5.** ______, The classification of simply connected manifolds of positive scalar curvature, Ann. Math., 111 (1980) 423–434.

**6.** R. S. Hamilton, Three-manifolds with positive Ricci curvature, J. Differential Geom., 17 (1982) 255–306.

**7.** W. Klingenberg, Lectures on Closed Geodesics, Grundlehren Math. Wiss. 230, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

**8.** S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, vol. I and II, Interscience Publishers, New York, 1963/69.

**9.** R. Schoen and S. T. Yau, On the structure of manifolds with positive scalar curvature, Manuscripta Math., 28 (1979) 159–183.

**10.** ______, Existence of incompressible minimal surfaces and the topology of manifolds with non-negative scalar curvature, Ann. Math., 110 (1979) 127–142.

**11.** ______, Complete three dimensional manifolds with positive Ricci curvature and scalar curvature, Seminar on Diff. Geom., S. T. Yau, ed., Ann. Math. Studies 102, Princeton Univ. Press, 1982.

**12.** M. Spivak, A Comprehensive Introduction to Differential Geometry, vols. I–V, Publish Or Perish, 1970/75.

**13.** S. T. Yau, Calabi's conjecture and some new results in algebraic geometry, Proc. Nat. Acad. Sci. U.S.A., 74 (1977) 1798–1799.

---

*Charles Babbage: Pioneer of the Computer.* By Anthony Hyman. Princeton University Press, Princeton, NJ, 1982. xiv + 287pp. + illustrations, $25.00.

JOHN T. O'DONNELL
*Department of Computer Science, Indiana University, Bloomington, IN 47405*

The modern era of computing began in the 1930's with the development of theoretical models of computation by Turing, Church, Kleene and others. World War II provided the impetus to

construct practical machines which led to the modern stored-program digital computers. The Difference Engines and Analytical Engines designed by Charles Babbage a century before stand as a fascinating prologue to the modern era: they embody many of the concepts of modern computers, but they were largely forgotten and had little influence.

The union of scientific and mathematical theory with engineering practice was the-central theme in Babbage's life. Although his mathematical training and early research concerned calculus and the theory of functions, he also spent great time and effort on learning engineering and industrial techniques, even going so far as learning to use power machine tools. For many years Babbage made it a practice to visit industries of all kinds in order to study both their engineering techniques and the economics of their operation, becoming England's leading expert on the application of science to manufacturing. This work culminated in Babbage's most influential book, *On the Economy of Machinery and Manufactures* (1832). Babbage recognized the social problems caused by the industrial revolution, but believed that the continued advance of science and engineering would enhance the quality of life, and that the real interests of management and labor coincided. A firm supporter of capitalism, and, after the revolutions of 1848, a foe of socialism, Babbage's economic writings influenced John Stuart Mill and Karl Marx.

Babbage's concern with practical applications of mathematics led directly to the Difference Engine. Appalled by the profusion of errors in mathematical tables compiled by hand, Babbage decided to build a reliable machine to calculate and automatically typeset tables using the method of finite differences. Although he stressed the importance of tables in navigation and the dire consequences of errors, Babbage was unable to convince politicians and bureaucrats that his idea was useful.* Over several years the government provided 17,000 pounds for the project, and Babbage also spent much of his own money, but more was needed and he completed only a small working prototype (which still works today). Hyman argues persuasively that only the obstacles of a dishonest chief engineer and government incomprehension prevented Babbage from successfully completing the Difference Engine.

In the process of simplifying and generalizing the Difference Engine, Babbage realized the importance of allowing the results of calculations to become inputs to further calculations. He described this form of self reference as "the engine eating its own tail". Of course the finite difference method does this, but always in a fixed pattern. A computation requiring a different pattern of calculations requires construction of a new machine, but a better solution is the construction of a machine whose pattern can be changed. This was exactly the problem faced by the French loom designers, leading to the famous Jacquard loom which allowed the weaver to specify a pattern by punching a set of holes in cards. Babbage adapted the Jacquard cards to his Analytical Engine.

Since the pattern of calculations is flexible, as in the Jacquard loom, the Analytical Engine must have a general ability to use arbitrary values as inputs to calculations, and to use the results in arbitrary ways. Therefore the Analytical Engine contained separate sections for representing numbers (the Store) and calculating (the Mill, in an analogy with the active part of a cotton mill). Using modern terminology, each number in the Store had a unique address. The Mill contained two number cylinders to hold the inputs to a calculation, and another to receive the result; these are called registers today. Arbitrary operations consisted of a simple sequence of steps: (1) move the first operand, specified by an address on a "variable card", into a Mill register; (2) move the second operand, specified by another variable card; (3) perform the arithmetic operation specified by an "operation card"; (4) move the result into the Store location specified by another variable card. This procedure is similar to the operation of "three address computers", except that the Analytical Engine had no real instructions containing an operation code and operands; the

*An example of the extreme conservatism of the nineteenth century scientific bureaucrats is the belief attributed by the Prime Minister to the Master of Trinity College, Cambridge, "that a century should pass before new discoveries in science are admitted into the course of academical instruction" [Hyman, p. 149].

operations and variable addresses came from separate card readers.

The "engine eating its own tail" suggested another advance, one of the most important of all: using the results of a calculation to affect the motions of the card readers, thereby controlling what the Engine will do next. Babbage hinged the cards together, making it possible for the card readers to back up to previous cards as well as going on to the next card. Some of the Engine's instructions could move cards through the readers without executing any operations punched on them, giving it the ability to perform "conditional go to statements". The Analytical Engine card readers are often compared with Hollerith cards used until recently in computers, but in fact they are more analogous to paper or magnetic tapes, which allow similar control.

The key operation in the Analytical Engine (as in computers) was the movement of data from place to place. One of Babbage's insights was to separate the mechanism that allowed data movement within the machine from the mechanism that controlled which movements actually occurred. The control mechanism contained a studded barrel with one position to control each piece of data movement hardware. An operation was then specified by putting studs in the barrel corresponding to the correct data movements. This approach simplified the machine and also made it possible to change the sequence of data movements executed for any instruction merely by inserting or removing studs. The alternative would be a total redesign of the control mechanism. This technique was rediscovered in the 1950's and is called microprogramming.

As Hyman points out, it is difficult to determine how close Babbage came to inventing some of the concepts of the modern computer. A computer design consists of a hierarchy of levels: logical organization; register transfer model; microcode host; logic gates; electronic implementation of gates. In contrast, Babbage's writings concentrate on the lowest level—gears, cams, linkages, torque, etc.—with little comprehensive documentation of higher-level issues. This appears to violate the currently popular idea of top-down design, but top-down design requires some understanding of the nature of the lower levels in order to design the higher levels. Babbage had no fixed low-level computer technology to work from.

Even some of the detailed mechanical descriptions contain hints of fundamentally important ideas. One example occurs in Henry Babbage's (son of Charles Babbage) description of the implementation of the "chain" [**3**, pp. 334–336], which corresponds roughly to the modern And gate. The chain transmitted a control signal to a destination only when all of a set of conditions were satisfied. Henry Babbage wrote of "another important principle largely adopted, viz., to break up every train of motion as far as possible into short courses, the last step of each furnishing a mere guide for a fresh start in the mechanism from the driving power". This might appear just to be a mechanical way to control losses through friction and to prevent the mechanism from jamming. However, it is similar to the form of logic signal amplification used in modern Very Large Scale Integrated Circuits, in which logic values (represented by a quantity of charge) must periodically be "amplified" by controlling the creation of a new charge from the power supply. Mead and Conway argue that such issues are fundamentally important to computation, and propose the term "physics of computational systems", in contrast to the conventional "mathematics of computation" [**2**, Chapt. 9].

The central theme in modern Computer Science is the use of abstraction and notation to deal with complexity. Both computers and programs may contain hundreds of thousands of components which individually have simple behaviors, but the analysis of their aggregate behavior can be extremely difficult. Therefore a crucial prerequisite to progress is the development of techniques for dealing with this complexity. Structured programming, hierarchical decomposition and microprogramming are examples of techniques which lessen the difficulty of designing hardware and software systems. Babbage did not organize the design of his Engines in levels of abstraction in the modern sense, but he recognized the problem: "The complicated relation which then arose amongst the various parts of the machinery would have baffled the most tenacious memory. I overcame that difficulty by improving and extending a language of signs, the Mechanical

Notation. ... By such means I succeeded in mastering trains of investigation so vast in extent that no length of years ever allotted to one individual could otherwise have enabled me to control." [**1**; quoted in Hyman, p. 165]. Babbage considered the Mechanical Notation one of his most important contributions, and computer scientists today would agree.

There are three striking differences between the Analytical Engines and computers: the implementation technology, the number system, and the representation of programs. The first two are actually insignificant. Babbage chose mechanical devices as his implementation technology for the same reason that modern designers use solid state electronics: in order to reach the best compromise among ease of manufacture, speed, reliability and cost. It was convenient to represent numbers on geared wheels in base 10, but it is more convenient to represent numbers with electrical charge or voltage in base 2. The binary number system is not essential to computers: if in the future the best available technology has three states, we will see base 3 computers. In fact, Babbage considered several other number systems. However, it is not clear whether Babbage considered representing programs in the Store instead of punched cards. The stored program concept is central to some theoretical results (such as the Halting Problem), and it was considered a major practical advance in the early years of electronic computers because it allowed the flexibility of self-modifying programs. For example, a program to increment all the elements of an array would contain in a loop a Load instruction addressing the first word of the array, an Increment instruction, a Store instruction addressing the first word, and then instructions to increment the address fields of the Load and Store instructions. Subsequent executions of the loop would then address consecutive elements of the array. The invention of index registers made this trick unnecessary, and advocates of structured programming consider self-modifying programs very bad. A new set of practical advantages to the stored program computer became evident in the 1950's: it made possible compilers and operating systems, which must treat programs as data.

Babbage wrote an extremely entertaining autobiography available in a modern edition [**1**] which partially describes the Engines, introduces his other research interests, and contains a vast store of anecdotes (he was much in demand by Society, and a good story teller). Hyman's biography gives a more systematic and complete description of Babbage's life and influence, and I highly recommend it. It is objective and scholarly but includes some trenchant comments (e.g., "Dionysus Lardner... ballooned across the engineering landscape of the time sustained by an inexhaustible supply of hot air." [p. 147]). Most importantly for the reader, Hyman relates Babbage's life to the contemporary scientific, political and social events. However, the book does not explain how the Engines worked. It contains several reproductions of technical drawings of the Engines, but not enough detail to follow their operation. Reprints of papers giving more technical details on the operation of the Engines are available in [**3**], and [**4**] reprints a wide range of papers on early computers.

Babbage met frustration in the construction of the Difference Engine, but he had confidence that both science and his ideas would triumph. According to Philip and Emily Morrison, "Babbage once said that he would gladly give up the remainder of his life if he could be allowed to live three days five hundred years hence and be provided with a scientific guide to explain the discoveries made since his death." [**3**, p. xxxi]. One hundred years later, computer science has justified his work, and its stream of discoveries shows no sign of slowing.

### References

**1.** Charles Babbage, Passages from the Life of a Philosopher, Longman, Green (1864). Facsimile reprint, Dawsons of Pall Mall, London, 1968.

**2.** Carver Mead and Lynn Conway, Introduction to VLSI Systems, Addison-Wesley, Reading, MA, 1980.

**3.** Philip Morrison and Emily Morrison (ed.), Charles Babbage and his Calculating Engines, Dover Publications, New York, 1961.

**4.** Brian Randell (ed.), The Origins of Digital Computers, Springer-Verlag, New York, 1975.

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

A. Wilansky (this MONTHLY, 91 (1984) 531) asks whether there is a normed $Q$-algebra which is not inverse closed in its completion. In fact a normed algebra is a $Q$-algebra if and only if it is inverse closed in its completion. Since the details are too technical for this letter column, interested readers may obtain a preprint from the undersigned.

Theodore W. Palmer
Department of Mathematics
University of Oregon
Eugene, OR 97403

---

Editor:

Recent articles in the MONTHLY by Renaud (March 1983, pp. 202–203) and Galvin (May 1984, pp. 308–309) address the question of generating integer matrices with known integer eigenvalues and eigenvectors. A very simple method for this, which has been used for many years (see, e.g., J. M. Ortega, Generation of Test Matrices by Similarity Transformations, Comm. ACM 7, 1964, pp. 377–78) by numerical analysts to generate test matrices is the following:

If $u$ and $v$ are column vectors with $u^Tv \neq -1$, then

$$(I + uv^T)^{-1} = I - \beta uv^T, \quad \beta = (1 + v^Tu)^{-1}.$$

If $D$ is a given integer diagonal matrix and $u$ and $v$ are integer vectors with $v^Tu = 0$, then

$$A = (I + uv^T)D(I + uv^T)^{-1} = D + u(Dv)^T - Duv^T - v^TDuuv^T$$

is an integer matrix with the same spectrum as $D$ and with eigenvectors that are the columns of $I + uv^T$. It is relatively easy to generate, either by hand for small matrices or computer program for larger ones.

Of course, the treatments by Renaud and Galvin have other strong points. However, for the purpose of generating a matrix with prescribed eigenvalues and known eigenvectors, the simplicity of the above approach has much to recommend it.

James M. Ortega
Department of Applied Mathematics
University of Virginia
Charlottesville, VA 22903

---

## 153. MISCELLANEA

A mathematician doesn't look like anything! Which means, he will always look so generally intelligent that there is no single definite thing behind it all!

—Robert Musil, *The Man Without Qualities*,
Coward-McCann, New York, 1953, p. 70.

# RAYMOND W. BRINK SELECTED PAPERS

**The Raymond W. Brink Selected Mathematical Papers Series** is a collection of papers on single topics selected and reprinted from the journals of the Mathematical Association of America. The papers are grouped by subject within the topic of the volume and are indexed by author. Each volume is a rich source of mathematical stimulation for students and teachers.

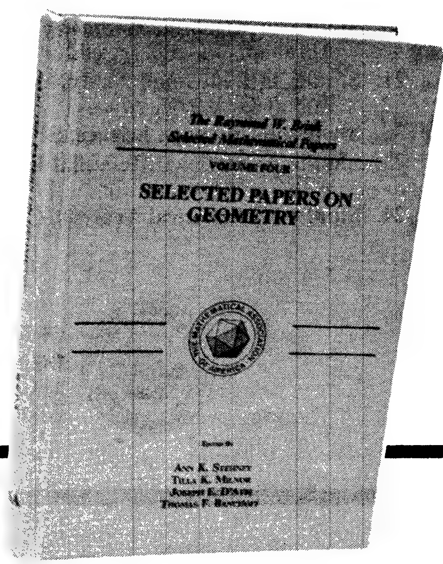
Four volumes are available in this series:

Volume 1.
**Selected Papers on Precalculus**
Edited by Tom M. Apostol, Gulbank D. Chakerian, Geraldine C. Darden, and John D. Neff.
xvii + 469 pages. Hardbound.
List: $24.00
MAA Members *(personal use):* $18.50

Volume 2.
**Selected Papers on Calculus**
Edited by Tom M. Apostol, Hubert E. Chrestenson, C. Stanley Ogilvy, Donald E. Richmond, and N. James Schoonmaker.
xv + 397 pages. Hardbound.
List: $24.00
MAA Members *(personal use):* $18.50

Volume 3.
**Selected Papers on Algebra**
Edited by Susan Montgomery, Elizabeth W. Ralston, S. Robert Gordon, Gerald J. Janusz, Murry M. Schacher, and Martha K. Smith.
xx + 537 pages. Hardbound.
List: $24.00
MAA Members *(personal use):* $18.50

Volume 4.
**Selected Papers on Geometry**
Edited by Ann K. Stehney, Tilla K. Milnor, Joseph D'Atri, and Thomas F. Banchoff.
ix + 338 pages. Hardbound.
List: $24.00
MAA Members *(personal use):* $18.50




Order From:
**The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

# From The Mathematical Association of America

**VOLUME 6—DOLCIANI MATHEMATICAL EXPOSITIONS**

## MAXIMA AND MINIMA WITHOUT CALCULUS,

by Ivan Niven
xv + 323 pp. Hardbound
List: $28.50 MAA Member: $21.30

Ivan Niven, distinguished author of several books on number theory and probability, has compiled the basic elementary techniques for solving maxima and minima problems. Since many books and courses already cover calculus and linear programming techniques, the author deliberately omits these areas from his discussions and concentrates instead on methods in algebra and geometry not so widely known. These methods are organized according to the mathematical ideas used, and many chapters can be read independently without reference to what precedes or follows. Some of the problems presented are left for the reader to solve with sketches of solutions given in the later pages.

The book is written for an audience at or near the maturity level of second- or third-year college students with a good working knowledge of precalculus mathematics. Although calculus is not a prerequisite, a prior knowledge of that subject will enhance the reader's comprehension. An excellent sourcebook in the area of maxima and minima, the book will serve as a textbook or as enrichment material for the talented undergraduate.

The main topics covered are:

- **Simple Algebraic Results**
- **Elementary Geometric Questions**
- **Isoperimetric Results**
- **Basic Trigonometric Inequalities**
- **Polygons Inscribed and Circumscribed**
- **Ellipses**
- **The Bees and Their Hexagons**
- **Further Geometric Results**
- **Applied and Miscellaneous Problems**
- **Euclidean Three-Space**
- **Isoperimetric Results not Assuming Existence**
- **Postscript on Calculus**

*Order from:*
**The Mathematical Association of America**
1529 Eighteenth St. NW
Washington, DC 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 92, Number 8 October 1985

# Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, Michigan State University, East Lansing, MI 48824. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS CO., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year





Second class postage paid at Washington, DC, and additional mailing offices

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# HERMITIAN MATRIX INEQUALITIES AND A CONJECTURE


N. N. CHAN

*Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*

MAN KAM KWONG*

*Department of Mathematics, Northern Illinois University, DeKalb, IL 60115*


**1. Introduction.** In many ways Hermitian matrices resemble real numbers. Indeed, all eigenvalues of a Hermitian matrix are real and the matrix is diagonalizable. This similitude may lead an unwary mind to wrong conclusions. This is especially true in the study of inequalities involving Hermitian matrices.

In the sequel we use capital letters $A, B, \ldots, X$, etc., to denote $n \times n$ Hermitian matrices where $n$ is some integer greater than 1; $A = A^*$ where $A^*$ denotes the conjugate of the transpose of $A$. We use $u$ and $v$ to denote complex column vectors in $C^n$ furnished with the usual inner product $(u, v)$. We define:

1. $A \geqslant (>) 0$ if all eigenvalues of $A$ are nonnegative (positive) or equivalently if $(u, Au) \geqslant (>) 0$ for all nonzero vectors $u \in C^n$.
2. $A \geqslant (>) B$ if $A - B \geqslant (>) 0$, or equivalently if $(u, Au) \geqslant (>)\ (u, Bu)$ for all nonzero vectors $u \in C^n$.

Unfortunately this ordering is only a partial one. Thus it is not true that if $A$ is not greater than or equal to $B$ then $A$ must be smaller than $B$. Another source of trouble is the fact that matrix multiplication is not compatible with the ordering (unless all matrices involved are mutually commutative). Most inequalities involving multiplications of real numbers cease to hold when the real numbers are replaced by Hermitian matrices. Two simple examples are:

1. It is not true that $A \geqslant 0$ and $B \geqslant 0$ imply that $AB + BA \geqslant 0$. Simply take $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.
2. It is not true that $A \geqslant B \geqslant 0$ implies that $A^2 \geqslant B^2$. Take $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

To illustrate the delicacy of Hermitian inequalities further, we note two more examples:

3. Let $A$ and $B$ be nonnegative Hermitian such that $A + B > 0$, and let $X$ be an $n \times n$ Hermitian matrix that satisfies the inequality

$$(A + B)X + X(A + B) \geqslant AB + BA. \tag{1.1}$$

In the scalar case ($n = 1$), it is obvious that the real number $X$ must be positive. However, that is

*N. N. Chan*: I received my bachelor's degree (1957) in probability theory from the University of Peking, and my Ph.D. (1963) in mathematical statistics under the supervision of R. L. Plackett and J. Aitchison from Liverpool University. I was lecturer in statistics at the University of Newcastle-upon-Tyne, 1963–1966, and currently am Senior Lecturer at the Chinese University of Hong Kong. Research interests: linear and nonlinear functional and structural relationships, majorization in $A$-optimal regression designs, screening of probability samples, order statistics, and Hermitian matrix inequalities. Leisure interest: table tennis.

*Man Kam Kwong*: I received my B.Sc. (1968) from the University of Hong Kong and my Ph.D. (1973) from the University of Chicago under the supervision of Professor Richard W. Beals. I have taught at Hong Kong Baptist College and Hong Kong Polytechnic. In 1977, I came to Northern Illinois University where I am currently professor. I spent a sabbatical leave at Argonne National Laboratory in 1982. Research interests: differential equations, matrix theory, functional analysis, numerical solution of differential equations. Leisure interest: travelling, microcomputers.


*This research was done while the second author was spending his sabbatical leave at Hong Kong Baptist College. He wishes to thank the College for its support and provisions of excellent working conditions.

not always true for $n \geqslant 2$. A counterexample is

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then $(A + B)X + X(A + B) = AB + BA$ but $X \not\geqslant 0$.

However, if in addition $A > 0$ and instead of (1.1) $X$ satisfies

$$(A + B)X + X(A + B) \geqslant A^{-1}B + BA^{-1}, \tag{1.2}$$

then $X \geqslant 0$. In fact if we have equality in (1.2), then $X$ is given by

$$X = A^{-1} - (A + B)^{-1} \geqslant 0.$$

Lemma 1 of Section 2 can be used to complete the proof for the general case. More generally, it can be proved that if $X$ satisfies

$$(A + B)X + X(A + B) \geqslant A^{\gamma}B + BA^{\gamma} \tag{1.3}$$

with $\gamma \in (-1, 0]$, then $X \geqslant 0$.

4. Let $A > 0$, $P \geqslant 0$ and let $t$ be a real number in $(-2, \infty)$. In the scalar case it is trivial that the solution of the following equation is nonnegative:

$$A^2X + XA^2 + tAXA = P \geqslant 0. \tag{1.4}$$

Can this be generalized to matrices? Equation (1.4) extends Lyapunov's Theorem (Lemma 1 of Section 2). Recently in **[10]** it was shown that the same result holds for $2 \times 2$ matrices. But for $3 \times 3$ matrices the conclusion is true only for $t \in (-2, 8]$. If $t > 8$, there is always a pair $(A, P)$ for which the solution $X$ is not nonnegative definite. For each dimension $n$, there is a cutoff value $t_0$ such that the result holds for $t \in (-2, t_0)$ but not for $t > t_0$. It is also known that $t_0 > 2$ for all $n$. The question of the exact value of $t_0$ for $n \geqslant 4$ is open.

Those few results concerning inequalities that continue to hold are usually hard to prove. The following is a natural question to ask:

What functions preserve the ordering of Hermitian matrices? In other words, what must $f$ satisfy so that

$$A \geqslant B \Rightarrow f(A) \geqslant f(B)?$$

Let us recall how functions of a Hermitian matrix are defined. If $D$ is a diagonal matrix with diagonal elements $\{d_1, d_2, \dots, d_n\}$, $a < d_i < b$, and $f\colon (a, b) \to (-\infty, \infty)$ is any real-valued function, then $f(D)$ is defined to be the diagonal matrix with diagonal elements $\{f(d_1), f(d_2), \dots, f(d_n)\}$. In general, if $A$ is not diagonal, there exist a unitary matrix $U$ and a diagonal matrix $D$ such that $A = U^*DU$. We then define $f(A) = U^*f(D)U$.

The question was studied and completely resolved by C. Loewner in his 1934 paper **[12]**. It turns out that the condition on $f$ depends on the size of the matrices in question. More precisely, let

$$P_n(a, b) = \{f\colon (a, b) \to (-\infty, \infty)\colon f(A) \geqslant f(B) \text{ for all pairs of } n \times n \text{ Hermitian matrices}$$
$$A, B \text{ with eigenvalues in } (a, b) \text{ such that } A \geqslant B\}.$$

Then

$$P_1(a, b) \supsetneqq P_2(a, b) \supsetneqq \cdots \supsetneqq P_n(a, b) \supsetneqq \cdots.$$

The class of functions $P_n(a, b)$ can be characterized by the positivity of a certain matrix formed with divided differences of the function with respect to $n$ arbitrary distinct points taken in $(a, b)$. It is necessary that $f \in P_n(a, b)$ possess continuous derivatives up to order $2n - 3$ and that $f^{(2n-3)}$ be convex.

Loewner went on to show that the set $P(a,b) = \cap_{n=2}^{\infty} P_n(a,b)$ is miraculously tied up with the theory of analytic functions. Indeed $P(a,b)$ is precisely the set of real-valued functions on $(a,b)$ that are continuable analytically to the upper half plane in which the functions take values with positive imaginary parts (the so-called Pick functions).

For an exposition of this beautiful theory, we refer the readers to the monograph by Donoghue [**6**], and the article by Ando [**1**].

From an advanced point of view, Loewner's result is elegant, but the proof is by no means short or elementary. For instance, it is not at all obvious that $f(t) = \sqrt{t}$ belongs to $P(0,\infty)$. Perhaps this is the reason why the fact "$t^r \in P(0,\infty), 0 < r < 1$" has been rediscovered more than once. We shall come back to this in the next section.

It is an interesting exercise to try to discover as many functions in $P(a,b)$ as possible by entirely elementary means. In Section 2 we look at this problem. In Sections 3 and 4 we discuss a conjecture and some results that are inspired by an attempt to look for further inequalities.

**2. An inequality-generating theorem.** We shall establish the following general result.

THEOREM 1. *Let $A, B, C, D$ be $n \times n$ Hermitian matrices. Suppose that $A$ commutes with $C$ and that $B$ commutes with $D$. If $A \geqslant B \geqslant 0$ and $C \geqslant D \geqslant 0$, then for any positive $r$ and $s$ such that $r + s \leqslant 1$ we have*

$$A^r C^s \geqslant B^r D^s.$$

The proof of Theorem 1 depends on the following very special case. (Take $r = 1/2$, $s = 0$, and $C = D = I$, the identity matrix).

THEOREM 2. *If $A \geqslant B \geqslant 0$, then $A^{1/2} \geqslant B^{1/2}$.*

As pointed out in the introduction, Theorem 2 is an obvious corollary of Loewner's deep result. R. Bellman [**3**] proved Theorem 2 analytically using ideas in the theory of differential equations. Y. H. Au-Yeung [**2**] gave a simple algebraic proof of the following generalization.

THEOREM 3. *If $A \geqslant B \geqslant 0$, then $A^r \geqslant B^r$ for all $r \in [0,1]$.*

Special cases of Theorem 1 were used by Au-Yeung in his proof of Theorem 3 but Theorem 1 has not been stated in such generality.

Theorem 3 was in fact "discovered" earlier in 1951 by E. Heinz [**7**], not just for matrices but for general Hermitian operators in a Hilbert space. In the following year T. Kato [**8**] gave a shorter proof. These authors seem not to have been aware of Loewner's result at the time the papers were written. The various proofs are, however, different and are worth recording. In [**9**] the second author gave yet another proof that resulted in a stronger form of Theorem 3.

Recently the operator version of Theorem 2 found an interesting application in linear neutron transport theory; see [**11**]. An important normequivalence result can now be given a simple elementary proof. This is an impetus for further investigation of related inequalities. Still more recently Theorem 2 was used in a study of the oscillation of second order systems of differential equations of the form $X''(t) + Q(t)X(t) = 0$, where $Q(t)$ is an $n \times n$ Hermitian matrix for all $t$ and $X(t)$ is an $n$-vector, see [**5**].

The following proof of Theorem 2 is a simplification of a proof that appeared in Marshall and Olkin [**13**]. Alternative proofs by other authors can also be found in [**13**].

*Proof of Theorem* 2. If suffices to establish the conclusion under the additional condition that $A > 0$, since the general case follows by continuity. Let

$$C = A^{1/2}, \quad D = B^{1/2} \quad \text{and} \quad X = C - D.$$

Then

$$0 \leqslant C^2 - D^2 = C^2 - (X - C)^2 = CX + XC - X^2.$$

Theorem 1 is thus a consequence of the following well-known result due to Lyapunov.

LEMMA 1. *Suppose $C$ and $P$ are square matrices of the same size with $C > 0$. The equation*

$$CX + XC = P \tag{2.1}$$

*has a unique solution $X$. Moreover if $P \geqslant 0$, then so is $X$.*

*Proof.* For the sake of completeness we give the proof. That (2.1) has a unique solution $X$ can be checked directly after we notice that we can assume without loss of generality that $C$ is diagonal. In particular, by taking adjoints of both sides of (2.1), we conclude that $X$ must be Hermitian whenever $P$ is. Suppose finally that $P \geqslant 0$ and $u$ is an eigenvector of $X$ corresponding to the eigenvalue $\alpha$. Then

$$0 \leqslant (u, Pu) = (u, CXu) + (u, XCu) = 2\alpha(u, Cu).$$

It follows that $\alpha \geqslant 0$ and hence $X$ is nonnegative.

An alternative proof is to make use of the explicit formula $X = \int_0^\infty e^{-Ct}Pe^{-Ct}\,dt$ which can be found in Bellman [**4**, p. 179].

The following lemmas can be found in [**2**] and [**9**].

LEMMA 2. *Suppose $C \geqslant 0$. Then $A \geqslant B$ implies that $CAC \geqslant CBC$. If furthermore $C > 0$, then the converse also holds.*

*Proof.* It is not difficult to prove the lemma using the definition of ordering via the inner product.

LEMMA 3. *If $A \geqslant B \geqslant 0$, then $B^{-1} \geqslant A^{-1}$.*

*Proof.* Let $I$ denote the identity matrix. First notice that $X \geqslant I$ implies that $X^{-1} \leqslant I$. This can be verified either directly from the definition of $X^{-1}$ or from Lemma 2 with $A = X$, $B = I$, and $C = X^{-1/2}$ Now Lemma 2 gives $B^{-1/2}AB^{-1/2} \geqslant I$ from the hypothesis that $A \geqslant B$. Taking the inverse we get $B^{1/2}A^{-1}B^{1/2} \leqslant I$. Then the conclusion follows after using Lemma 2 one more time.

LEMMA 4. *Suppose that $G \geqslant F > 0$ and $X \geqslant 0, Y \geqslant 0$. If $YFY \geqslant XGX$, then $Y \geqslant X$.*

*Proof.* By Lemma 2, $YGY \geqslant YFY$, which together with the second hypothesis gives $YGY \geqslant XGX$. It follows from Lemma 2 that

$$G^{1/2}YGYG^{1/2} \geqslant G^{1/2}XGXG^{1/2} \quad \text{or} \quad (G^{1/2}YG^{1/2})^2 \geqslant (G^{1/2}XG^{1/2})^2.$$

By Theorem 2, $G^{1/2}YG^{1/2} \geqslant G^{1/2}XG^{1/2}$. Lemma 2 then gives the conclusion.

We are now ready to give a proof of Theorem 1.

*Proof of Theorem* 1. As in the proof of Theorem 2, we may assume without loss of generality that $B > 0$. By Lemma 3, $B^{-1} \geqslant A^{-1}$.

Using Lemma 4 with $X = (BD)^{1/2}$, $Y = (AC)^{1/2}$, $F = A^{-1}$ and $G = B^{-1}$, we obtain the conclusion of the Theorem for the special case $r = 1/2$ and $s = 1 - r$.

Replacing $A$ and $B$ by $(AC)^{1/2}$ and $(BD)^{1/2}$, respectively, in the special case $r = 1/2$ of the Theorem, we obtain the special case $r = 1/4$, $s = 3/4$.

Replacing $C$ and $D$ by $(AC)^{1/2}$ and $(BD)^{1/2}$, respectively, in the special case $r = 1/2$, we obtain the special case $r = 3/4$, $s = 1/4$. Repeating this process, we see that the Theorem holds when $r$ is any dyadic fraction, namely, when $r$ is of the form $k/2^n$ ($n = 1, 2, \ldots; k = 1, 2, \ldots, 2^n - 1$) and $s = 1 - r$. Since such fractions are dense in $[0, 1]$, the Theorem holds for all general $r \in (0, 1)$ and $s = 1 - r$ by continuity. For $s < 1 - r$, the required inequality can be written as

$$(A^{r'}C^{s'})^{r+s} \geqslant (B^{r'}D^{s'})^{r+s}, \text{where } r' = r/(r+s),\ s' = s/(r+s) \quad \text{and} \quad r' + s' = 1.$$

We can thus complete the proof of Theorem 1 by using the result established above and Theorem 3, which as has been pointed out is a special case of Theorem 1 when $C = D = I$, $r + s = 1$.

An immediate consequence of Theorem 1 is the following property of the Pick functions. Let

$P_0(0,\infty)$ be the subfamily

$$\{f \in P(0,\infty)\colon f\colon (0,\infty) \to [0,\infty)\} \text{ of } P(0,\infty).$$

THEOREM 4. *For any $r, s > 0$ such that $r + s \leqslant 1$ and any $f, g \in P_0(0,\infty)$, we have $f^r g^s \in P_0(0,\infty)$. More generally, let $r_i > 0$ $(i = 1,2,\ldots,n)$ be such that $\sum_{k=1}^{m} r_i \leqslant 1$. Then $\prod_{i=1}^{m} f_i^{r_i} \in P_0(0,\infty)$ for all $f_i \in P_0(0,\infty)$ $(i = 1,2,\ldots,n)$.*

In other words, the set of functions $\{\ln f\colon f \in P_0(0,\infty)\}$ is convex.

Let us show how further examples of functions in $P_0(0,\infty)$ can be generated. We assume throughout the hypotheses $A \geqslant B \geqslant 0$; in a few cases the stronger condition $B \geqslant 0$ is needed. Notice that in any case we need only establish the result under the stronger condition $B > 0$ as the general case follows by continuity. Also notice that Theorem 4 can be used in the obvious way to yield generalizations.

Let $r, s > 0, r + s \leqslant 1$ and $\alpha, \beta > 0$. Then a simple consequence of Theorem 4 is

$$(1) \qquad (\alpha I + A)^r(\beta I + A)^s \geqslant (\alpha I + B)^r(\beta I + B)^s.$$

For all $\alpha \geqslant \gamma > 0, 0 \leqslant r \leqslant 1$

$$(2) \qquad (\gamma I + A)(\alpha I + A)^{-r} \geqslant (\gamma I + B)(\alpha I + B)^{-r},$$

or equivalently for all $\beta > 0$, $\beta^{-1} \geqslant \gamma$

$$(3) \qquad (\gamma I + A)(I + \beta A)^{-r} \geqslant (\gamma I + B)(I + \beta B)^{-r}.$$

Indeed by Lemma 3, $A^{-1} \leqslant B^{-1}$. Thus $(I + \alpha A^{-1}) \leqslant (I + \alpha B^{-1})$. Lemma 3 again yields $(I + \alpha A^{-1})^{-1} \geqslant (I + \alpha B^{-1})^{-1}$ which is exactly (2) with $r = 1$, $\gamma = 0$. Inequality (2) with $\gamma = 0$ now follows if in Theorem 1 we let $C = A(\alpha I + A)^{-1}$ and $D = B(\alpha I + B)^{-1}$. To see that (2) holds in the general case $\gamma \neq 0$ we replace $A$ by $\gamma I + A$ and $\alpha$ by $\alpha - \gamma$.

Integrating a known inequality with respect to some parameter leads to another inequality. The following are three examples.

$$(4) \qquad A[\ln(I + A)]^{-r} \geqslant B[\ln(I + B)]^{-r}$$

for all $r \in [0,1]$. We require $B > 0$ to guarantee that $[\ln(I + B)]^{-r}$ is defined. However, in the singular case we can still define $B[\ln(I + B)]^{-r}$ by a limiting process, since $\lim_{t \to 0} t[\ln(1 + t)]^{-r}$ exists as can be seen using L'Hospital's rule. By Theorem 3,

$$(I + A)^r \geqslant (I + B)^r, \qquad r \in [0,1].$$

Integrating this inequality with respect to $r$ over $[0,1]$ gives (4) with $r = 1$. The general case then follows upon applying Theorem 1.

In (2) we let $\gamma = 0$ and $r = 1$ and integrate the resulting inequality with respect to $\alpha$ over $[0,1]$ and finally invoke Theorem 1 to get

$$(5) \qquad A\left[\ln(I + A^{-1})\right]^r \geqslant B\left[\ln(I + B^{-1})\right]^r.$$

On the other hand, if in (3) we let $r = 1$, restrict $\gamma$ to be in [0,1], and integrate with respect to $\beta$ over [0,1], we obtain

$$(6) \qquad (I + \gamma A^{-1})\ln(I + A) \geqslant (I + \gamma B^{-1})\ln(I + B).$$

In particular we have

$$(7) \qquad \ln(I + A) \geqslant \ln(I + B).$$

We let the readers investigate for themselves what will result from integrating, for instance, $rA^r \geqslant rB^r$, $r \in [0,1]$, or inequality (2) with $\gamma = 0$, $r = 1$ and $\alpha$ replaced by $\alpha^2$, $\alpha \in [0,b]$.

In (7) we can replace $A$ by $nA$ and $B$ by $nB$ to deduce

$$\ln\left(\frac{1}{n} + A\right) \geqslant \ln\left(\frac{1}{n} + B\right).$$

Letting $n \to \infty$ we get

$$\ln A \geqslant \ln B. \tag{8}$$

Here $f(t) = \ln t$ is an example of a function belonging to $P(0,\infty)$ but not to $P_0(0,\infty)$. Inequality (8) can also be derived from Theorem 3 via differentiation. By Theorem 3,

$$A^r - I \geqslant B^r - I.$$

Dividing by $r$ and taking limits as $r \to 0$, we have

$$\left.\frac{dA^r}{dr}\right|_{r=0} \geqslant \left.\frac{dB^r}{dr}\right|_{r=0},$$

which is (8).

The findings above can be summarized as follows:

Examples of functions in $P_0(0,\infty)$ include

| | | |
|---|---|---|
| $(\alpha+t)^r(\beta+t)^s$ | $\alpha,\beta>0$ | $r,s>0, r+s\leqslant 1$ |
| $(\gamma+t)/(\alpha+t)^r$ | $\alpha\geqslant\gamma>0$ | $r\in[0,1]$ |
| $(\gamma+t)/(1+\beta t)^r$ | $\beta^{-1}\geqslant\gamma>0$ | $r\in[0,1]$ |
| $t/\ln^r(1+t)$ | $r\in[0,1]$ | |
| $t\ln^r(1+1/t)$ | $r\in[0,1]$ | |
| $(1+\gamma/t)\ln(1+t)$ | $\gamma\in[0,1]$ | |
| $(t\ln t - t + 1)\ln^{-2} t$ | | |
| $\sqrt{t}\,\tan^{-1}(b/\sqrt{t})$ | $b\in(0,\infty)$; | |

$\ln t$ belongs to $P(0,\infty)$. Both $P_0(0,\infty)$ and $P(0,\infty)$ are closed cones (hence the integration technique for generating new examples works) and furthermore $P_0(0,\infty)$ satisfies Theorem 4. It should be stated again that all these can be directly deduced from Loewner's original result.

**3. A conjecture.** Taking square roots of course restores the inequality that is destroyed after squaring. It is interesting to ask if the following sequence of operations will preserve the inequality:

1. Squaring the two sides of the inequality $A \geqslant B \geqslant 0$ to get $(A^2, B^2)$.
2. Multiplying both on the right and on the left by some $C > 0$ to get $(CA^2C, CB^2C)$.
3. Taking square roots.

The question is whether

$$(CA^2C)^{1/2} \geqslant (CB^2C)^{1/2}.$$

Lemma 2 seems to indicate that the second operation should not have worsened the situation. Thus one is tempted to answer yes.

However, after some trial and error using the MAT-LAB package on the VAX at Argonne National Laboratory, the following counterexample was discovered. Take

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 101 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix},$$

and

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 0.2 \end{pmatrix}.$$

It is easy to check that $A - B \geqslant 0$. The computer output shows (originally with 16 decimal

places)

$$E = \left(CA^2C\right)^{1/2} - \left(CB^2C\right)^{1/2} = \begin{pmatrix} 4.116681\ldots & 1.679158\ldots \\ 1.679158\ldots & 0.134061\ldots \end{pmatrix}.$$

The eigenvalues of $E$ are $-0.4794\ldots$ and $4.730\ldots$. Thus $(CA^2C)^{1/2} \not\geqslant (CB^2C)^{1/2}$.

The same experiment seems to indicate that with the special choice $C = B$ or $A$, the inequality persists. We therefore propose the following conjecture. We have found a proof for the case of the lowest dimension $n = 2$. It is proved by sheer brute force although the proof is by no means straightforward.

CONJECTURE. *If $A \geqslant B \geqslant 0$, then*

$$(BA^2B)^{1/2} \geqslant B^2, \tag{3.1}$$

*and*

$$A^2 \geqslant (AB^2A)^{1/2}. \tag{3.2}$$

Notice that the second inequality is a consequence of the first one. We may assume without loss of generality that $B > 0$ so that $B^{-1}$ exists. By hypothesis, $B^{-1} \geqslant A^{-1} > 0$. The first inequality then gives $(A^{-1}B^{-2}A^{-1})^{1/2} \geqslant A^{-2}$. Taking inverses now gives (3.2).

It is tempting to guess that the similar inequality

$$(AB^2A)^{1/2} \geqslant B^2$$

also follows from the hypotheses of the conjecture. If this were true, then by transitivity (using (3.2) at least for $n = 2$), it would follow that $A^2 \geqslant B^2$, which we know is false.

Repeated use of (3.1) shows that

$$(B^3A^2B^3)^{1/2} \geqslant B^4, \tag{3.3}$$

and more generally

$$(B^{m-1}A^2B^{m-1})^{1/2} \geqslant B^m \tag{3.4}$$

for $m = 2^k$, where $k = 1, 2, \ldots$. We know that (3.4) is true at least in the two-dimensional case. It is only natural to ask if (3.4) is true for all $m \in (1, \infty)$ and for all dimension $n$.

**4. Some evidence for the conjecture.** Let us see how Conjecture 1 leads to another proposition. Let $B$ be any positive Hermitian matrix and $P$ be any nonnegative matrix. Define

$$Y(t) = \left[B(B + tP)^2 B\right]^{1/2}.$$

By Conjecture 1, $Y(t) \geqslant Y(0)$ for $t \geqslant 0$. Thus

$$Y'(0) \geqslant 0. \tag{4.1}$$

By differentiating the equation $Y^2(t) = B(B + tP)^2B$ and then letting $t = 0$, we see that $X = Y'(0)$ satisfies the matrix equation

$$B^2X + XB^2 = B^2PB + BPB^2. \tag{4.2}$$

Notice that $Q = BPB$ is also nonnegative. Thus the validity of the following is a consequence of that of Conjecture 1.

THEOREM 5. *Let $B$ be a positive Hermitian matrix and $Q$ a nonnegative Hermitian matrix. The solution $X$ of the following matrix equation is always nonnegative*:

$$B^2X + XB^2 = BQ + QB. \tag{4.3}$$

If the right-hand side is nonnegative, then, by Lemma 1, the conclusion holds. However, as noted before $BQ + QB$ is in general not nonnegative.

*Proof of Theorem* 5. Let $Y$ be the solution of

$$B^2Y + YB^2 = Q. \tag{4.4}$$

By Lemma 1 of Section 2, $Y \geqslant 0$. Thus it follows from Lemma 2 that $BYB \geqslant 0$.

Direct computation shows that $BY + YB$ satisfies equation (4.3). Hence by uniqueness (Lemma 1)

$$X = BY + YB.$$

Now

$$\begin{aligned} BX + XB &= B(BY + YB) + (BY + YB)B \\ &= Q + 2BYB \\ &\geqslant 0. \end{aligned}$$

Thus by Lemma 1 again $X \geqslant 0$.

COROLLARY 6. *Let $b_1, \dots, b_n$ be $n$ positive numbers. The determinant*

$$\det\left(\frac{b_i b_j (b_i^2 + b_j^2)}{(b_i^4 + b_j^4)}\right)_{i,j=1,\dots,n} \tag{4.5}$$

*is nonnegative.*

*Proof.* Apply Theorem 5 with $B = \operatorname{diag}(b_1, \dots, b_n)$ and $Q = B^{1/2}RB^{1/2}$ where all the entries of $R$ are 1—direct computation shows that the solution of (4.3) is given by (4.5).

It was originally planned to derive Theorem 5 as a consequence of Corollary 6 which can be proved directly in case $n = 2$ and 3.

It is interesting to point out that the following generalization of Theorem 5 is still true.

THEOREM 7. *Let $s$ be any real number in $[1, \infty)$ and $B, Q$ be as in Theorem 5. Then the solution of the matrix equation*

$$B^sZ + ZB^s = BQ + QB \tag{4.6}$$

*is nonnegative.*

Direct computation easily confirms the case $n = 2$. Lemma 1 can be thought of as the limiting case $s \to \infty$ (with $B^s = C$ and $B = C^{1/s} \to$ identity). Notice that the Theorem is false for $s < 1$. Theorem 7 and other generalizations will be given in a forthcoming paper.

A most natural question to ask is: How can one characterize all functions $f$ such that the solution of the matrix equation $f(B)Z + Zf(B) = BQ + QB$ is nonnegative?

To conclude the paper we mention another tantalizing conjecture.

PROBLEM. *Let $A$, $B$ and $C$ be nonnegative Hermitian matrices such that $A \leqslant C, B \leqslant C$. Is it true that*

$$(A^2 + B^2)^{1/2} \leqslant \sqrt{2}\,C?$$

There are counterexamples to the more general conjecture: $A \leqslant C, B \leqslant D$ implies $(A^2 + B^2)^{1/2} \leqslant (C^2 + D^2)^{1/2}$.

**References**


**1.** T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, Linear Algebra and Appl., 26 (1979) 203–241.

**2.** Y. H. Au-Yeung, Some inequalities for the rational power of a nonnegative definite matrix, Linear Algebra Appl., 7 (1973) 347–350.

**3.** R. E. Bellman, Some inequalities for the square root of a positive definite matrix, Linear Algebra Appl., 1 (1968) 321–324.

**4.** ______, Introduction to Matrix Analysis, McGraw-Hill, New York, 1970.

5. R. Byers, B. J. Harris, and Man Kam Kwong, Weighted means and oscillation conditions for second order matrix differential equations, J. Differential Equations (to appear).

6. W. F. Donoghue, Monotone Matrix Functions and Analytic Continuation, Springer-Verlag, Berlin, 1974.

7. E. Heinz, Beitrage zur Storungsteorie der Spektralzerlegung, Math. Ann., 123 (1951) 415–538.

8. T. Kato, Notes on some inequalities for linear operators, Math. Ann., 125 (1952) 208–212.

9. Man Kam Kwong, Inequalities for the powers of nonnegative Hermitian operators, Proc. Amer. Math. Soc., 51 (1975) 401–406.

10. ______, On the definiteness of the solution of certain matrix equation, Linear Algebra and Appl. (to appear).

11. Man Kam Kwong and H. Kaper, preprint.

12. C. Loewner, Uber monotone Matrixfunktionen, Math. Z., 38 (1934) 177–216.

13. A. W. Marshall and I. Olkin, Inequalities: Theory of Majorization and Its Applications, Academic Press, New York, 1979.

# MONOSTATIC SIMPLEXES*

R. J. MacG. DAWSON

*Department of Mathematics, Corpus Christi College, Cambridge, England CB2 1RH*

It is a fact, learned by many of us at an early age, that certain objects are liable to fall over if stood the wrong way up. Flat-faced objects are usually considered to be safer, but it is easy to design a polyhedron that is unstable on some of its faces. Some years ago, it was shown that a polyhedron may be constructed which is *monostatic* (unstable on all but one of its faces) [**1**].

In the same article, it was stated that no tetrahedron can be monostatic, although the published proof was incomplete, making no explicit use of the position of the center of mass of a tetrahedron. (That this is essential may be seen from the construction, due to Conway, of a *non-uniform* tetrahedron with a center of mass very near one corner which is stable only on one face. James Tappin has suggested to the author the picturesque term *vanka-vstanka* for a non-uniform polytope stable on only one face, from the Russian word for a doll that rights itself when knocked over.) A complete proof, not published, was given by Conway [**1**], [**2**], who also asked various related questions. This paper presents a partial answer to one of these: in how few dimensions does a monostatic simplex exist? While I cannot give a sharp answer to this, it will be shown that no such simplex exists in six or fewer dimensions, and that there is an example in ten. Additional references and open problems are given in [**4**].

**1. Nonexistence of Monostatic Simplexes in Low Dimensions.** The nonexistence result is essentially a more quantitative version of the following

THEOREM (Conway). *No tetrahedron is monostatic.*

*Proof.* Obviously, a simplex cannot tip about an edge unless the dihedral angle at that edge is obtuse. As the altitude, and hence the height of the barycenter, is inversely proportional to the area of the base for any given tetrahedron, a tetrahedron can only tip from a smaller face to a larger one.

Suppose some tetrahedron to be monostatic, and let $A$ and $B$ be the largest and second-largest faces respectively. Either the tetrahedron rolls from another face, $C$, onto $B$ and thence onto $A$, or

---

*R. J. MacG. Dawson:* I was born in 1961, in Halifax, Nova Scotia, into an academic but definitely non-mathematical family. My B. Sc. (mathematics and physics), is from the University of King's College, in Halifax, although all my courses were at Dalhousie University. (Any King's student or alumnus will explain this situation to you if you buy him or her a drink.) I am currently working in geometry under Dr. Hallard Croft, in the hope of eventually getting my Ph.D. My outside interests include cycling, pastel drawing, playing the recorder, and hiking; and I have recently bought a motorbike and am learning to ride it.

*Work was supported by an 1851 Exhibition Scholarship.

else it rolls from $B$ to $A$ and also from $C$ to $A$. In either case, one of the two largest faces has two obtuse dihedral angles, and one of them is on an edge shared with the other of the two largest faces.

The projection of the remaining face, $D$, onto the face with two obtuse dihedral angles must be as large as the sum of the projections of the other three faces. But this makes the area of $D$ larger than that of the face we are projecting onto, contradicting our assumption that $A$ and $B$ are the two largest faces. ■

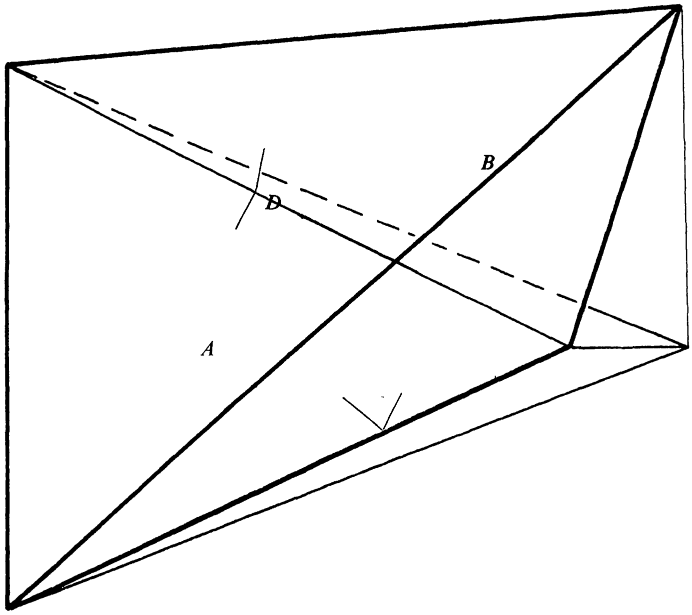



FIG. 1. If the dihedral angles shown are obtuse, face $A$ cannot be larger than face $D$.

In more than three dimensions, it will be convenient to use a representation for the simplexes based on a theorem of Minkowski. A set of vectors in $\mathbb{R}^n$ is said to be *equilibrated* if the following hold:

(1a) $$\sum \mathbf{x}_i = 0,$$

(1b) $$\{\mathbf{x}_i\} \text{ spans } \mathbb{R}^n,$$

(1c) $$\mathbf{x}_i \neq a\mathbf{x}_j \quad \text{for any } i \neq j,\ a \geqslant 0.$$

For any face of a polytope, the *face vector* is the vector with length equal to the area of the face and directed normal to the face.

THEOREM (Minkowski). *Every equilibrated set of vectors in $\mathbb{R}^n$ is the set of face vectors for a unique convex polytope.*

*The set of face vectors of any convex polytope is equilibrated.*

(For proof, see **[5]**; there is a discussion of this result in **[3]**.)

We will refer to the set of face vectors as the *Minkowski representation* of the polytope. In this representation, there is a particularly simple criterion to determine whether a simplex will tip between two given faces.

PROJECTION CRITERION. *Let, for each face $F_i$ of a simplex, $\mathbf{x}_i$ be the face vector, and let $\theta_{ij}$ be the angle between $\mathbf{x}_i$ and $\mathbf{x}_j$. The simplex can tip from $F_i$ to $F_j$ if and only if*

(2) $$\|\mathbf{x}_i\| < \|\mathbf{x}_j\| \cos \theta_{ij}.$$

*Proof.* Let $\mathbf{v}_i$ be the vertex opposite $F_i$. The barycenter $B$ is given vectorially by:

$$\mathbf{B} = \frac{\sum \mathbf{v}_i}{n+1}. \tag{3}$$

For any vector $\mathbf{x}$, let $p_1(\mathbf{x})$ be the projection of $\mathbf{x}$ into the plane $(F_i \cap F_j)^{\perp}$. Note that $p_1(\mathbf{v}_i)$ is the altitude of $F_j$ from $(F_i \cap F_j)$; hence

$$\frac{\|\mathbf{x}_i\|}{\|\mathbf{x}_j\|} = \frac{\|p_1(\mathbf{v}_j)\|}{\|p_1(\mathbf{v}_i)\|}. \tag{4}$$

Project again into the line $(F_i \cap F_j)^{\perp} \cap F_i$;

$$\begin{aligned} p_2(\mathbf{B}) &= \frac{p_2(\mathbf{v}_i) + p_2(\mathbf{v}_j)}{n+1} \\ &= \frac{\|p_1(\mathbf{v}_j)\| - \|p_1(\mathbf{v}_i)\|\cos\theta_{ij}}{n+1}. \end{aligned} \tag{5}$$

The simplex tips from $F_i$ to $F_j$ if and only if

$$p_2(\mathbf{B}) < 0,$$

$$\Leftrightarrow \|p_1(\mathbf{v}_j)\| < \|p_1(\mathbf{v}_i)\|\cos\theta_{ij},$$

$$\Leftrightarrow \|\mathbf{x}_i\| < \|\mathbf{x}_j\|\cos\theta_{ij}. \quad \blacksquare \tag{2}$$

LEMMA. *Suppose $F_m, F_{m-1}, \dots, F_0$ are faces of a simplex which tips from $F_i$ to $F_{i-1}$ for $i = 1, 2, \dots, m$. Then*

$$\|\mathbf{x}_m\|\cos\theta_{0m} \geqslant -\left(\cos\left(\frac{\pi}{m+1}\right)\right)^{m+1}\|\mathbf{x}_0\|. \tag{6}$$

*Proof.* By the projection criterion (2),

$$\frac{\|x_i\|}{\|x_{i-1}\|} < \cos\theta_{i\,i-1};$$

so

$$\frac{\|x_m\|}{\|x_0\|} < \prod_{i=1}^{m}\cos\theta_{i\,i-1},$$

$$\frac{\|x_m\|\cos\theta_{0m}}{\|x_0\|} > \min\left(0, \cos\theta_{0m}\prod_{i=1}^{m}\cos\theta_{i\,i-1}\right). \tag{7}$$

By the triangle inequality,

$$\sum_{i=1}^{m}\theta_{i\,i-1} \geqslant \theta_{0m},$$

and hence

$$(\pi - \theta_{0m}) + \sum_{i=1}^{m}\theta_{i\,i-1} \geqslant \pi. \tag{8}$$

Minimizing the right hand side of (7) under the constraint (8), we see that the left member of (7) is greater than the minimum of $-\prod_{i=0}^{m}\cos\alpha_i$ taken over sets of $\alpha_i$ whose sum is $\pi$, and this immediately gives (6). $\blacksquare$

THEOREM. *No simplex in $\mathbb{R}^6$ is monostatic.*

*Proof.* Label the faces in decreasing order of size (i.e., with $F_0$ the largest face and with $F_6$ the

smallest). We consider two cases, depending on whether the simplex tips onto $F_0$ from one face or from more than one face.

*Case* 1. $F_1$ is the only face from which the simplex tips onto $F_0$.

If we place the simplex on $F_i$, $i > 1$, it reaches $F_1$ after at most $(i-1)$ falls. Thus, by the lemma,

$$\frac{\|\mathbf{x}_i\|\cos\theta_{1i}}{\|\mathbf{x}_1\|} > -\cos^i\left(\frac{\pi}{i}\right). \tag{9}$$

Using this and the other inequalities

$$\frac{\|\mathbf{x}_0\|\cos\theta_{01}}{\|\mathbf{x}_1\|} > 1, \tag{10}$$

$$\begin{aligned} \frac{\|\mathbf{x}_2\|\cos\theta_{12} + \|\mathbf{x}_6\|\cos\theta_{16}}{\|\mathbf{x}_1\|} &\geqslant \frac{\|\mathbf{x}_2\|\cos\theta_{12} - \|\mathbf{x}_6\|}{\|\mathbf{x}_1\|} \\ &> \frac{\|\mathbf{x}_2\|}{\|\mathbf{x}_1\|}(\cos\theta_{12} - 1) \\ &> \cos^2\theta_{12} - \cos\theta_{12} \\ &\geqslant -\tfrac{1}{4}, \end{aligned} \tag{11}$$

we can derive the following lower bound on the projection of the sum of the face vectors onto $\mathbf{x}_1$:

$$\begin{aligned} \sum_{i=0}^{6}\frac{\|\mathbf{x}_i\|\cos\theta_{1i}}{\|\mathbf{x}_1\|} &> \frac{7}{4} - \cos^3\frac{\pi}{3} - \cos^4\frac{\pi}{4} - \cos^5\frac{\pi}{5} \\ &= (77 - 5\sqrt{5})/64 \\ &> 0, \end{aligned} \tag{12}$$

contradicting our assumption that the vectors form an equilibrated set.

*Case* 2. There are two or more penultimate faces.

We deal similarly with this case, using the bounds

$$\frac{\|\mathbf{x}_1\|\cos\theta_{01} + \|\mathbf{x}_6\|\cos\theta_{06}}{\|\mathbf{x}_0\|} > -\frac{1}{4} \quad \text{for } F_1 \text{ and } F_6, \tag{11$'$}$$

$$\frac{\|\mathbf{x}_i\|\cos\theta_{0i}}{\|\mathbf{x}_0\|} > 0 \quad \text{for any penultimate face other than } F_1, \tag{13}$$

and (6) for any other face, given that the simplex rolls from $F_i$ to $F_0$ in at most $(i-1)$ rolls. Thus:

$$\begin{aligned} \sum_{i=0}^{6}\frac{\|\mathbf{x}_i\|\cos\theta_{0i}}{\|\mathbf{x}_0\|} &> \frac{3}{4} - \cos^3\frac{\pi}{3} - \cos^4\frac{\pi}{4} - \cos^5\frac{\pi}{5} \\ &= (13 - 5\sqrt{5})/64 \\ &> 0. \quad \blacksquare \end{aligned} \tag{14}$$

**2. Existence of a Monostatic Simplex in $\mathbb{R}^{10}$.** Suppose a set of vectors in $\mathbb{R}^m$ to be equilibrated; if we embed the $\mathbb{R}^m$ in a higher-dimensional $\mathbb{R}^n$, (1a) and (1c) will remain satisfied. If the set has more than $n$ elements, we may also perturb it slightly so that it spans $\mathbb{R}^n$ without losing the other properties. If there are $n+1$ vectors, the perturbed vectors must be the Minkowski representation of a simplex; and if, in the original set of vectors, for every $\mathbf{x}_i$ but one

there exists an $\mathbf{x}_j$ satisfying (2), then for small enough perturbations the resulting simplex will be monostatic, as (2) will continue to hold in each case.

Let $\theta_i$, $i = 1,2,3,4,5$, be defined as follows:

$$\theta_1 = \frac{81\pi}{227}, \quad \theta_2 = \frac{54\pi}{227}, \quad \theta_3 = \frac{36\pi}{227}, \quad \theta_4 = \frac{24\pi}{227}, \quad \theta_5 = \frac{16\pi}{227},$$

and define, for any fixed $\varepsilon > 0$, $\{\mathbf{x}_0, \mathbf{x}_{\pm i} | i = 1,2,3,4,5\}$:

$$\arg(\mathbf{x}_i) = -\arg(\mathbf{x}_{-i}) = \sum_{j=1}^{i} \theta_j, \arg(\mathbf{x}_0) = 0 \tag{15a}$$

$$\|\mathbf{x}_i\| = \|\mathbf{x}_{-i}\| = \left(\prod_{j=1}^{i} \cos\theta_j\right) - \varepsilon, \tag{15b}$$

$$\|\mathbf{x}_0\| = -\sum_{i=1}^{5}\left[(\|\mathbf{x}_i\| + \|\mathbf{x}_{-i}\|)\cos\sum_{j=1}^{i}\theta_j\right]. \tag{15c}$$

It follows from (15c) that the vectors are equilibrated. From (15a) and (15b), we see that for $i = 1, 2, 3$ or $4$ and any small positive $\varepsilon$ the projection criterion (2) is satisfied for a fall from $F_{i+1}$ to $F_i$, or from $F_{-(i+1)}$ to $F_{-i}$. It remains only to verify (2) for a fall from $F_{\pm 1}$ to $F_0$. As $\|\mathbf{x}_{\pm 1}\| = \cos\theta_1 - \varepsilon$, it is sufficient to show that, for small enough $\varepsilon$, $\|\mathbf{x}_0\| > 1$. In fact, if we calculate the right hand side of (15c) numerically as $\varepsilon$ approaches zero in (15b), we obtain a value of $1.21953+$, and hence (2) is satisfied for small enough $\varepsilon$ in (15b). The relative sizes and angles of these vectors are shown in Fig. 2. (The $3:2$ ratio between successive angles was found heuristically; it does not in fact give the set of $\theta_i$ which maximizes $\|\mathbf{x}_0\|$, but comes tantalizingly close! It would be interesting to know why this happens.)

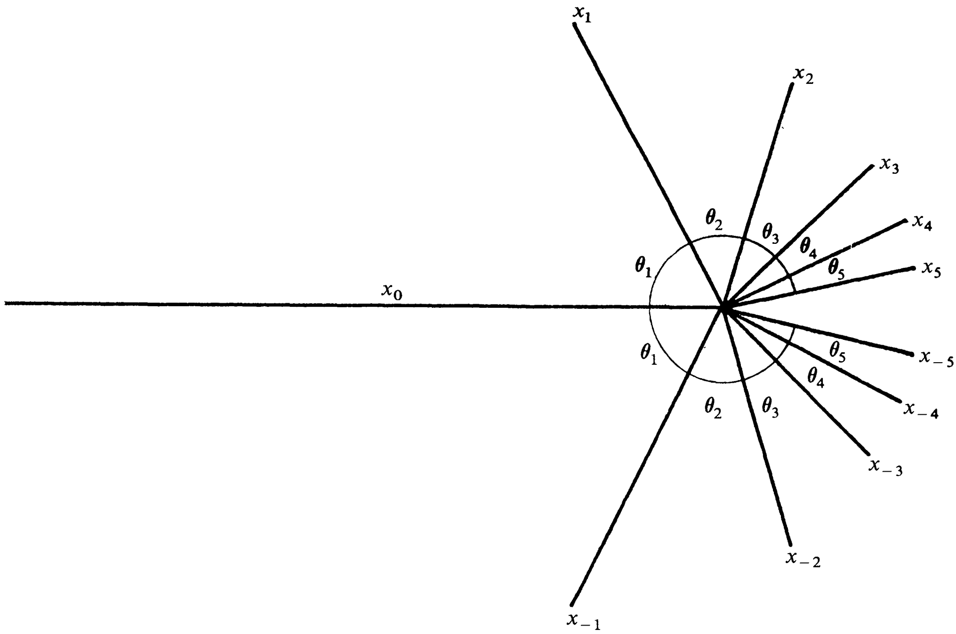



**FIG. 2. Plane vectors approximating the face vectors of a ten-dimensional monostatic simplex.**

**Acknowledgement.** I should like to thank Professor J. H. Conway for several suggestions improving the clarity of this paper.

**References**

1. J. H. Conway and R. K. Guy, Stability of Polyhedra, SIAM Rev., 11 (1969) 78–82.
2. J. H. Conway, private communication, 1983.
3. B. Grünbaum, Convex Polytopes, Pure and Applied Mathematics, Vol. 16, New York, 1967, 339–340.
4. R. K. Guy, Twenty odd questions in combinatorics, in "Combinatorial Mathematics and its Applications," R. C. Bose, et al., eds., Chapel Hill, NC, 1970, pp. 209–232.
5. H. Minkowski, Allgemeine Lehrsätze Über die Konvexe Polyeder, Nachr. Ges. Wiss. Göttingen (1897), 198–219.

# THE ISOPERIMETRIC POINT AND THE POINT(S) OF EQUAL DETOUR IN A TRIANGLE

G. R. VELDKAMP
*Biltsteijn 2, 3732 GT De Bilt, The Netherlands*

This paper takes its being from an asterisk occurring in Problem E 3020, proposed by Clark Kimberling, University of Evansville. It runs as follows:

Suppose $ABC$ is a nonisosceles triangle. Find three hyperbolas concurrent in a point $P$ such that triangles $APB$, $APC$, and $BPC$ all have the same perimeter. (*) How does this common perimeter compare with that of $ABC$?

As usual in this Journal the asterisk indicates that at the time of publication neither the proposer nor the editors of the problem section were in possession of a solution.

It is not difficult to describe unambiguously the three hyperbolas referred to in the problem, but one discovers at once that these curves are not very informative as to a full understanding of the problem. Several questions arising in this connection have been completely treated in the present paper.

For readers not very well acquainted with some geometric terms occurring in the paper References [1] and [2] might come in useful.

**1. The problem.** A finite point $X$ in the plane of the triangle $ABC$ will be called an *isoperimetric point of this triangle* if the triangles $XBC$, $XCA$ and $XAB$ all have the same perimeter $2\sigma$. This paper deals with the following questions:

Has any triangle an isoperimetric point? If not, find a necessary and sufficient condition for a triangle to have such a point (§§7, 8). Is this point related to well-known remarkable points of the triangle? How could one construct it by ruler and compasses? Does there exist an upper and lower bound for the ratio $\sigma/s$, where $2s$ stands for the perimeter of $\Delta ABC$? (§10). It will be shown that answers to these questions can be obtained by analytic as well as synthetic means.

**2. Notations and some frequently used formulas.** The sideline of $\Delta A_1A_2A_3$ opposite $A_k$ will be denoted by $l_k$, the side (and its length) lying on $l_k$ by $a_k$ and the angle opposite $a_k$ by $\alpha_k$. Furthermore, $I$ and $r$ stand for the center and the radius respectively of the inscribed circle $\gamma_0$; $B_k$ is the point of contact of $\gamma_0$ with $a_k$. The radius of the escribed circle of $\Delta A_1A_2A_3$ having its center on $A_kI$ will be noted by $r_k$. For briefness' sake we shall use the abbreviations $\sin\frac{1}{2}\alpha_k = s_k$, $\cos\frac{1}{2}\alpha_k = c_k$ and $\tan\frac{1}{2}\alpha_k = t_k$ $(k = 1,2,3)$. Finally, $2s$, $F$ and $R$ denote in this order the

Dr. G. R. Veldkamp is an emeritus professor at the University of Technology, Eindhoven (The Netherlands) where he was teaching mathematics and kinematics. After his retirement he lives at leisure with his wife in a small village in the center of his country. His favorite occupations consist of reading and collecting first-class detective-novels and solving not too hard mathematical problems.

perimeter, the area and the radius of the circumcircle of $\Delta A_1A_2A_3$. The following formulas and the formulas obtained by cyclic permutation shall be used frequently:

$$s_2c_3 + s_3c_2 = c_1; c_2c_3 - s_2s_3 = s_1; s_2c_2 + s_3c_3 = c_1(c_2c_3 + s_2s_3); \tag{2.1}$$

$$r = F/s; r_1 = F/(s - a_1); R = a_1a_2a_3/4F. \tag{2.2}$$

Furthermore:

$$s = 4Rc_1c_2c_3; r = 4Rs_1s_2s_3; r_k = st_k \qquad (k = 1,2,3). \tag{2.3}$$

Using the suitable formulas from the collection, we find:

$$s(t_1 + t_2 + t_3) = 4Rc_1c_2c_3(t_1 + t_2 + t_3) = 4R(s_1c_2c_3 + s_2c_3c_1 + s_3c_1c_2) = 4R(1 + s_1s_2s_3);$$

hence:

$$r_1 + r_2 + r_3 = 4R + r. \tag{2.4}$$

We observe finally that

$$t_2t_3 + t_3t_1 + t_1t_2 = 1. \tag{2.5}$$

Except for §8, we assume throughout that $\Delta A_1A_2A_3$ is nonisosceles.

**3. Three hyperbolas.** Let $P$ be a point in the plane of $\Delta A_1A_2A_3$. We put $\overline{PA_k} = p_k$ $(k = 1,2,3)$. By definition $P$ is an isoperimetric point of $\Delta A_1A_2A_3$ if and only if

$$p_2 + a_1 + p_3 = p_3 + a_2 + p_1 = p_1 + a_3 + p_2 = 2\sigma. \tag{3.1}$$

This is equivalent to

$$p_3 - p_1 = a_3 - a_1 \quad \text{and} \quad p_1 - p_2 = a_1 - a_2. \tag{3.2}$$

Let $A_k^*$ be the mirror image of $A_k$ into the perpendicular bisector of $a_k$ $(k = 1,2,3)$. The first of the equations (3.2) expresses the fact that $P$ is a point of the branch passing through $A_2^*$ of the hyperbola $h_2$ going through $A_2$ and having $A_1$ and $A_3$ as its real foci. The second equation (3.2) states that $P$ is on the branch through $A_3^*$ of the hyperbola $h_3$ going through $A_3$ with $A_1$ and $A_2$ as its real foci. It follows from (3.2) that $p_3 - p_2 = a_3 - a_2$. This means that any common point of the aforementioned branches also lies on the branch going through $A_1^*$ of the hyperbola $h_1$ with $A_2$ and $A_3$ as its foci and passing through $A_1$.

Our conclusion is: *an isoperimetric point of a triangle is a common point of three well-defined branches of three uniquely determined hyperbolas.*

**4. The equations of $h_2$ and $h_3$.** We shall use homogeneous triangular point coordinates $x_1, x_2, x_3$ with respect to the basic points $A_1(1,0,0)$, $A_2(0,1,0)$, $A_3(0,0,1)$ and the unit point $I(1,1,1)$.

Let $X_k$ be the orthogonal projection of $X(x_1, x_2, x_3)$ onto $l_k$ and $d_k$ the directed distance from $X$ to $l_k$; then

$$d_k = fx_k(k = 1,2,3; f \neq 0)$$

and

$$2F = a_1d_1 + a_2d_2 + a_3d_3 = f(a_1x_1 + a_2x_2 + a_3x_3).$$

Hence,

$$f = \frac{2F}{a_1x_1 + a_2x_2 + a_3x_3}. \tag{4.1}$$

We have in the quadrangle $A_1X_3XX_2$ (Fig. 1) the relation: $\overline{XA_1}\sin\alpha_1 = \overline{X_2X_3}$ and therefore

$$\overline{XA}_1^2\sin^2\alpha_1 = \overline{X_2X_3}^2 = d_2^2 + d_3^2 + 2d_2d_3\cos\alpha_1 = f^2(x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1).$$

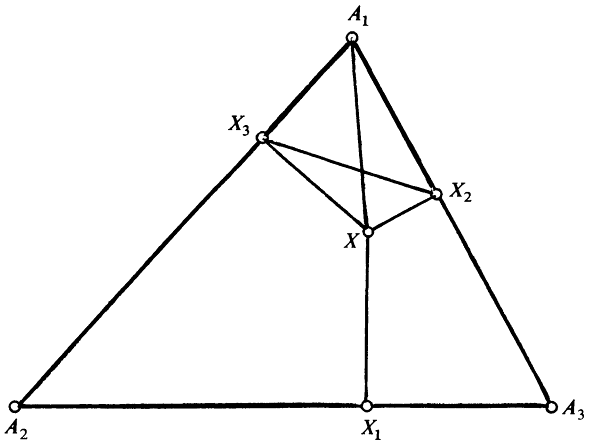



FIG. 1

Consequently,

$$\overline{XA_1}\sin\alpha_1 = |f|\left(x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1\right)^{1/2} \tag{4.2}$$

The locus of the points with zero distance to $A_1$ is the pair of isotropic lines through $A_1$. These lines are therefore represented by $x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1 = 0$. Since $A_1$ is a focus of $h_2$, the said lines are tangents of $h_2$. If $u_1x_1 + u_2x_2 + u_3x_3 = 0$ represents the join of the points of contact (i.e., the directrix of $h_2$ corresponding to $A_1$), the equation of $h_2$ takes the form

$$\left(u_1x_1 + u_2x_2 + u_3x_3\right)^2 - \lambda\left(x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1\right) = 0. \tag{4.3}$$

Because $h_2$ passes through $A_2$ and has the internal bisector $x_1 - x_3 = 0$ of $\angle A_1A_2A_3$ as a tangent, we must have $\lambda = u_2^2$ and

$$u_1 - u_2\cos\alpha_1 + u_3 = 0. \tag{4.4}$$

The directrices of $h_2$ are perpendicular to the principal axis $x_2 = 0$, that is, parallel to the altitude $x_1\cos\alpha_1 - x_3\cos\alpha_3 = 0$ through $A_2$. Since the line at infinity is represented by

$$x_1\sin\alpha_1 + x_2\sin\alpha_2 + x_3\sin\alpha_3 = 0,$$

we must therefore have

$$\begin{vmatrix} \cos\alpha_1 & 0 & -\cos\alpha_3 \\ \sin\alpha_1 & \sin\alpha_2 & \sin\alpha_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = 0,$$

or

$$u_1\cos\alpha_3 - u_2 + u_3\cos\alpha_1 = 0. \tag{4.5}$$

It follows from (4.4) and (4.5) that $u_1$, $u_2$ and $u_3$ are proportional to $\sin^2\alpha_1$, $\cos\alpha_3 - \cos\alpha_1$ and $\cos\alpha_1\cos\alpha_3 - 1$, respectively. From these facts results the following equation for $h_2$:

$$\left\{x_1\sin^2\alpha_1 + x_2(\cos\alpha_3 - \cos\alpha_1) + x_3(\cos\alpha_1\cos\alpha_3 - 1)\right\}^2$$
$$-(\cos\alpha_3 - \cos\alpha_1)^2\left(x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1\right) = 0.$$

We know that $A_1$ is also a focus of $h_3$. This permits us to obtain the equation of $h_3$ simply by interchanging the indices 2 and 3 in the equation for $h_2$. This yields

$$\{ x_1 \sin^2\alpha_1 + x_2(\cos\alpha_1\cos\alpha_2 - 1) + x_3(\cos\alpha_2 - \cos\alpha_1)\}^2$$
$$-(\cos\alpha_2 - \cos\alpha_1)^2(x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1) = 0$$

as the equation for $h_3$.

**5. Pairs of lines in the three pencils determined by $h_1, h_2$ and $h_3$ taken two by two.** Elimination of $x_2^2 + x_3^2 + 2x_2x_3\cos\alpha_1$ between the equation of $h_2$ and $h_3$ yields

$$\begin{aligned}(\cos\alpha_2 - \cos\alpha_1)^2\{ x_1\sin^2\alpha_1 + x_2(\cos\alpha_3 - \cos\alpha_1) + x_3(\cos\alpha_1\cos\alpha_3 - 1)\}^2 \\ = (\cos\alpha_3 - \cos\alpha_1)^2\{ x_1\sin^2\alpha_1 + x_2(\cos\alpha_1\cos\alpha_2 - 1) + x_3(\cos\alpha_2 - \cos\alpha_1)\}^2.\end{aligned} \tag{5.1}$$

This equation obviously represents two straight lines constituting a degenerate conic of the pencil $\Pi_{23}$ determined by the pair $\{h_2, h_3\}$. An easy computation reveals that (5.1) is the equation of the pair of lines $\{l, n_{23}\}$, where $l$ is represented by

$$c_1^2(c_2^2 - c_3^2)x_1 + c_2^2(c_3^2 - c_1^2)x_2 + c_3^2(c_1^2 - c_2^2)x_3 = 0, \tag{5.2}$$

and $n_{23}$ by

$$s_1^2(c_2^2 + c_3^2 - 2c_1^2)x_1 + s_2^2(c_1^2 - c_3^2)x_2 + s_3^2(c_1^2 - c_2^2)x_3 = 0. \tag{5.3}$$

The equation of $l$ is invariant under cyclic permutation of the indices 1, 2, 3, which cannot be said of equation (5.3). We infer that the pencils $\Pi_{31}$ and $\Pi_{12}$ determined by the pairs $\{h_3, h_1\}$ and $\{h_1, h_2\}$ contain the degenerate conics $\{l, n_{31}\}$ and $\{l, n_{12}\}$, respectively, where $n_{31}$ and $n_{12}$ are represented by

$$s_1^2(c_2^2 - c_3^2)x_1 + s_2^2(c_3^2 + c_1^2 - 2c_2^2)x_2 + s_3^2(c_2^2 - c_1^2)x_3 = 0 \tag{5.4}$$

and

$$s_1^2(c_3^2 - c_2^2)x_1 + s_2^2(c_3^2 - c_1^2)x_2 + s_3^2(c_1^2 + c_2^2 - 2c_3^2)x_3 = 0. \tag{5.5}$$

This proves

THEOREM 1. *The hyperbolas $h_1, h_2$ and $h_3$ have two points in common lying on $l$.*

We observe that in the equation of $l$ the cofactors of $x_1$, $x_2$ and $x_3$ add up to zero and that the same is true for the cofactors of $c_1^2x_1$, $c_2^2x_2$ and $c_3^2x_3$. The meaning of this is that the point $I(1,1,1)$ and the point $G(c_2^2c_3^2, c_3^2c_1^2, c_1^2c_2^2)$ are on $l$. It is easily seen that $G$ is the point of concurrence of the lines $A_1B_1$, $A_2B_2$ and $A_3B_3$; it is generally known as the *Gergonne point* of $\Delta A_1A_2A_3$. Hence the

COROLLARY. *The line $l$ of Theorem 1 is the join of the center of the inscribed circle to the Gergonne point of the triangle.*

In the equations (5.3), (5.4) and (5.5), the sum of the cofactors of $s_1^2x_1$, $s_2^2x_2$ and $s_3^2x_3$ vanishes. This means that $n_{23}$, $n_{31}$ and $n_{12}$ have the point $N(s_2^2s_3^2, s_3^2s_1^2, s_1^2s_2^2)$ in common. It is the isotomic conjugate of $G$ and generally known as *Nagel's point.*

**6. The coordinates of the points common to $h_1$, $h_2$ and $h_3$.** To find these coordinates, we determine the intersections of $GI$ with $h_3$. We shall use for our aim a somewhat modified equation of $h_3$ which can be obtained by the following argument. We observe that

$$\overline{B_3A_1} - \overline{B_3A_2} = s - a_1 - (s - a_2) = a_2 - a_1 = \overline{A_3A_1} - \overline{A_3A_2}.$$

Therefore $B_3(c_2^2, c_1^2, 0)$ is on $h_3$ and the same is true for the isotomic conjugate $B_3'(s_2^2, s_1^2, 0)$. Hence $h_3$ is a member of the pencil of conics passing through $A_3, B_3, B_3'$ and having $x_1 - x_2 = 0$ as their tangent at the point $A_3$. This pencil is represented by

$$(c_1^2x_1 - c_2^2x_2)(s_1^2x_1 - s_2^2x_2) + \lambda x_3(x_1 - x_2) = 0. \tag{6.1}$$

Since $IB_3 \perp A_1A_2$, the line $IB_3$ touches $h_3$ at $B_3$. Hence $A_3B_3$ is the polar line of $I$ with respect to $h_3$. The equation of $A_3B_3$ is $c_1^2x_1 - c_2^2x_2 = 0$, whereas the polar line of $I(1,1,1)$ with respect to (6.1) is given by

$$(c_1^2 - c_2^2)(s_1^2x_1 - s_2^2x_2) + (s_1^2 - s_2^2)(c_1^2x_1 - c_2^2x_2) + \lambda(x_1 - x_2) = 0.$$

Hence

$$(c_1^2 - c_2^2)(s_1^2c_2^2 - s_2^2c_1^2) = \lambda(c_1^2 - c_2^2)$$

and $\lambda = c_2^2 - c_1^2$. The equation of $h_3$ is therefore

$$(c_1^2x_1 - c_2^2x_2)(s_1^2x_1 - s_2^2x_2) + (c_2^2 - c_1^2)x_3(x_1 - x_2) = 0. \tag{6.2}$$

By means of a parameter $\mu$ any point on $GI$ may be represented by

$$(c_2^2c_3^2 + \mu, c_3^2c_1^2 + \mu, c_1^2c_2^2 + \mu).$$

Substituting these coordinates into (6.2) we find that this point is on $h_3$ if and only if $\mu^2 = c_1^2c_2^2c_3^2$. This proves

THEOREM 2. *The common points of $h_1$, $h_2$ and $h_3$ are*

$$Q(c_2c_3(c_2c_3 + c_1), c_3c_1(c_3c_1 + c_2), c_1c_2(c_1c_2 + c_3)) \tag{6.3}$$

*and*

$$P(c_2c_3(c_2c_3 - c_1), c_3c_1(c_3c_1 - c_2), c_1c_2(c_1c_2 - c_3)). \tag{6.4}$$

*With $I := (c_1c_2c_3, c_1c_2c_3, c_1c_2c_3)$ we have in symbolic notation*

$$Q = G + I, P = G - I.$$

Hence the

COROLLARY. *The two common points of the three hyperbolas are harmonic conjugates with respect to G and I.*

**7. The position of $P$ and $Q$ on $h_3$.** To decide which of the points $P$ and $Q$ is on which of the branches of $h_3$, we introduce the point $Z_\varepsilon$ with coordinates

$$(c_2^2c_3^2 + \varepsilon c_1c_2c_3, c_3^2c_1^2 + \varepsilon c_1c_2c_3, c_1^2c_2^2 + \varepsilon c_1c_2c_3),$$

where $\varepsilon^2 = 1$. Hence $Z_\varepsilon$ coincides with $P$ or $Q$ according as $\varepsilon = -1$ or $\varepsilon = 1$. We put moreover $\overline{QA_k} = q_k$ and $\overline{Z_\varepsilon A_k} = z_k$ $(k = 1,2,3)$ and we proceed to compute $z_k$ by means of (4.1) and (4.2). To find the value of $f_\varepsilon$ of $f$ corresponding to $Z_\varepsilon$ we have to evaluate

$$\begin{aligned} &a_1(c_2^2c_3^2 + \varepsilon c_1c_2c_3) + a_2(c_3^2c_1^2 + \varepsilon c_1c_2c_3) + a_3(c_1^2c_2^2 + \varepsilon c_1c_2c_3) \\ &\quad = 4Rc_1c_2c_3(s_1c_2c_3 + s_2c_3c_1 + s_3c_1c_2) + 2\varepsilon s c_1c_2c_3. \end{aligned}$$

Using the formulas from §2, we find that this is equal to $c_1c_2c_3(4R + r + 2\varepsilon s)$. Hence

$$f_\varepsilon = \frac{2F}{c_1c_2c_3(4R + r + 2\varepsilon s)} = \frac{32R^2s_1s_2s_3}{4R + r + 2\varepsilon s}. \tag{7.1}$$

It may be observed that our result shows that $Q$ is always a finite point, whereas $P$ is at infinity if and only if $4R + r = 2s$. We suppose further on that $4R + r \neq 2s$. By means of (4.2) we obtain

$$\begin{aligned} 2s_1c_1z_1 = |f_\varepsilon| \Big\{ &(c_3^2c_1^2 + \varepsilon c_1c_2c_3)^2 + (c_1^2c_2^2 + \varepsilon c_1c_2c_3)^2 \\ &+ 2(2c_1^2 - 1)(c_3^2c_1^2 + \varepsilon c_1c_2c_3)(c_1^2c_2^2 + \varepsilon c_1c_2c_3) \Big\}^{1/2} \end{aligned}$$

This leaves us with

$$z_1 = \frac{1}{2}|f_\varepsilon|s_1^{-1}c_1\left(c_2^2 + c_3^2 + \varepsilon c_1c_2c_3\right), \tag{7.2}$$

and by cyclic permutation

$$z_2 = \frac{1}{2}|f_\varepsilon|s_2^{-1}c_2\left(c_3^2 + c_1^2 + 2\varepsilon c_1c_2c_3\right) \quad \text{and} \quad z_3 = \frac{1}{2}|f_\varepsilon|s_3^{-1}c_3\left(c_1^2 + c_2^2 + 2\varepsilon c_1c_2c_3\right). \tag{7.3}$$

It follows that

$$z_1 - z_2 = \frac{1}{2}|f_\varepsilon|s_1^{-1}s_2^{-1}\left\{(s_2c_1 - s_1c_2)\left(c_3^2 + 2\varepsilon c_1c_2c_3\right) + c_1c_2\left(s_2c_2 - s_1c_1\right)\right\},$$

or

$$z_1 - z_2 = \frac{1}{2}|f_\varepsilon|s_1^{-1}s_2^{-1}(s_2c_1 - s_1c_2)(1 + s_1s_2s_3 + 2\varepsilon c_1c_2c_3).$$

Using (7.1) we arrive at

$$z_1 - z_2 = \frac{4Rs_3(s_2c_1 - s_1c_2)(4R + r + 2\varepsilon s)}{|4R + r + 2\varepsilon s|} = (a_2 - a_1)\frac{4R + r + 2\varepsilon s}{|4R + r + 2\varepsilon s|}. \tag{7.4}$$

Taking $\varepsilon = 1$ we get: $q_1 - q_2 = a_2 - a_1$. This means: *$Q$ is on the branch of $h_3$ passing through $A_3$.* If we take $\varepsilon = -1$ we are dealing with $P$ and we have to distinguish two cases:

i. $4R + r > 2s$. Then $p_1 - p_2 = a_2 - a_1$, which means that $P$ is on the branch passing through $A_3$.

ii. $2s > 4R + r$. In this case $P$ lies on the branch of $h_3$ passing through $A_3^*$. We therefore find as a final result of this paragraph:

THEOREM 3. *A triangle has an isoperimetric point if and only if* $2s > 4R + r$.

It is known from elementary geometry that a triangle is acute-, right- or obtuse-angled according as $s > 2R + r$, $s = 2R + r$ or $s < 2R + r$. Hence the

COROLLARY. *A triangle without an isoperimetric point is always obtuse-angled.*

Fig. 2 gives an idea of the situation for a triangle with an isoperimetric point $P$.

**8. Transformation of the condition $2s > 4R + r$.** It will be clear that it is not possible to decide at first sight whether a given triangle fulfills the condition of Theorem 3. In this paragraph we replace this condition therefore by an equivalent one that is somewhat more transparent.

We have $4R + r = r_1 + r_2 + r_3 = s(t_1 + t_2 + t_3)$; the inequality of Theorem 3 is therefore equivalent to $t_1 + t_2 + t_3 < 2$. Let us suppose $\alpha_3 > \alpha_2 \geqslant \alpha_1$, i.e., $t_3 > t_2 \geqslant t_1$. From (2.5) we get

$$t_3 = (1 - t_1t_2)/(t_1 + t_2)$$

and therefore

$$t_1 + t_2 + t_3 = \left(t_1^2 + t_1t_2 + t_2^2 + 1\right)/(t_1 + t_2).$$

The system of the preceding inequalities is therefore equivalent to

$$(t_1 + t_2 - 1)^2 - t_1t_2 < 0 \wedge t_2^2 + 2t_1t_2 - 1 < 0 \wedge t_2 \geqslant t_1 > 0. \tag{8.1}$$

We consider $(t_1, t_2)$ as a point $T$ with rectangular coordinates $t_1$ and $t_2$. The equation $(x + y - 1)^2 - xy = 0$ represents an ellipse $\varepsilon$ touching the coordinate-axes at $A(1,0)$ and $B(0,1)$ (Fig. 3) and meeting the line $x = y$ at $C(1,1)$ and $D(\frac{1}{3}, \frac{1}{3})$. The equation $y^2 + 2xy - 1 < 0$ is that of a hyperbola $h$ with asymptotes $y = 0$ and $2x + y = 0$; it meets $\varepsilon$ at $B$ and $E(4/3, 1/3)$ and the line $x = y$ at $F(\sqrt{3}/3, \sqrt{3}/3)$ and $F'(-\sqrt{3}/3, -\sqrt{3}/3)$. The inequalities (8.1) are

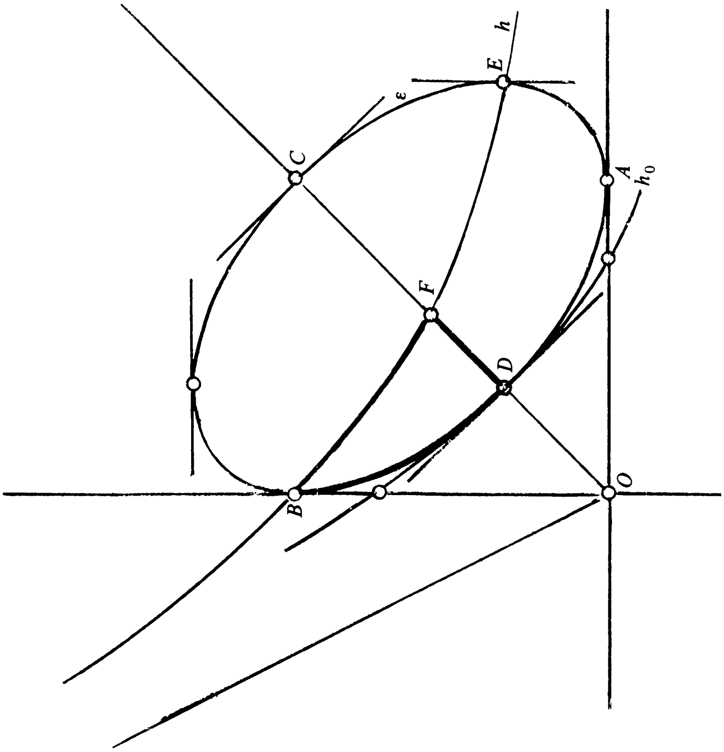



FIG. 3

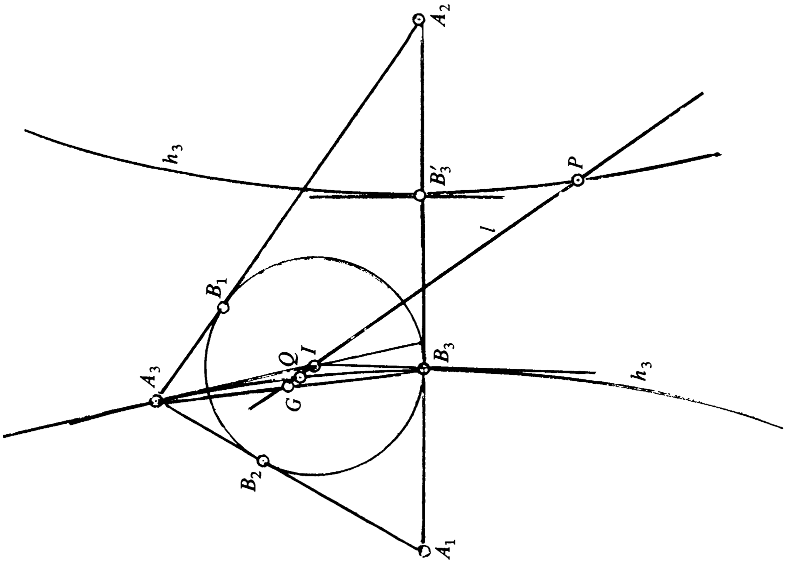



FIG. 2

geometrically equivalent to the assertion that $F(t_1, t_2)$ lies either in the interior of the curvilinear triangle $DFB$ shown in Fig. 3 or on the line $x = y$ between $D$ and $F$. We introduce the pencil of conics given by $xy + \lambda(x + y) - 1 = 0$, where $\lambda$ is a parameter. Its members are equilateral hyperbolas having only two real points in common, namely the points at infinity of the coordinate-axes. Only those parts of these hyperbolas are of interest for our problem, which are lying in the first quadrant. Clearly these parts meet the line $x = y$ between $D$ and $F$ if and only if they go through a point $F(t_1, t_2)$ in the interior of the curvilinear triangle $DFB$. A member of the pencil on hand goes through $F$ if and only if

$$\lambda = (1 - t_1 t_2)/(t_1 + t_2) = t_3.$$

If $(u, u)$ is its intersection with $x = y$, we have therefore

$$t_3 = (1 - u^2)/2u,$$

where $\frac{1}{3} < u < \frac{1}{3}\sqrt{3}$. Since $(1 - u^2)/2u$ is monotonic decreasing on this interval, we infer $\frac{1}{3}\sqrt{3} < t_3 < \frac{4}{3}$; hence $\pi/3 < \alpha_3 < 2$ arctan $4/3$. The first inequality is trivial because $\alpha_3$ was assumed to be the largest angle of $\Delta A_1 A_2 A_3$. The second half of the inequality provides us with a rather practical condition for a triangle to have an isoperimetric point, namely

THEOREM 4. *A necessary and sufficient condition for a triangle to have an isoperimetric point is that its largest angle not exceed* 2 arcsin $4/5 \approx 106°15'37''$.

It may be observed that the equilateral hyperbola $h_0$ of the aforementioned pencil passing through $D$ corresponds to an isosceles triangle with 2 arcsin $4/5$ as its vertical angle; it does not have an isoperimetric point.

**9. An expression for $\sigma$ in terms of $R$, $r$ and $s$.** If $P$ is the isoperimetric point of $\Delta A_1 A_2 A_3$, we find by means of (7.2) and (7.3) that

$$p_1 + p_2 + p_3 = \frac{1}{2}|f|(s_1 s_2 s_3)^{-1}\{s_2 s_3 c_1(c_2^2 + c_3^2 - 2c_1c_2c_3) + s_3 s_1 c_2(c_3^2 + c_1^2 - 2c_1c_2c_3)$$
$$+ s_1 s_2 c_3(c_1^2 + c_2^2 - 2c_1c_2c_3)\}, \tag{9.1}$$

where

$$|f| = \frac{2F}{c_1c_2c_3(2s - 4R - r)}. \tag{9.2}$$

Formula (9.1) can be simplified to

$$p_1 + p_2 + p_3 = \frac{1}{2}|f|(s_1 s_2 s_3)^{-1} c_1c_2c_3(1 + 4s_1s_2s_3 - 2c_1c_2c_3).$$

Using (9.2) we get, in view of $F = rs = 4Rss_1s_2s_3$,

$$p_1 + p_2 + p_3 = \frac{(4R + 4r - 2s)s}{2s - 4R - r}.$$

From (3.1) we derive $p_1 + p_2 + p_3 + s = 3\sigma$. Hence

$$\sigma = \frac{rs}{2s - 4R - r}. \tag{9.3}$$

**10. Inequalities for the ratio $\sigma/s$.** It will be clear from the above that $\sigma/s$ has no finite upper bound. The inequalities

$$p_2 + p_3 > a_1, p_3 + p_1 > a_2 \quad \text{and} \quad p_1 + p_2 > a_3$$

are valid for any triangle and any position of $P$. Hence $p_1 + p_2 + p_3 > s$. Combining this with $3\sigma = p_1 + p_2 + p_3 + s$, we find $\sigma/s > 2/3$. This inequality can be improved by observing that the minimum value of $\sigma/s$ corresponds to the maximum of $(2s - 4R - r)/r$. Using the

inequality

$$s - 2R \leqslant (3\sqrt{3} - 4)r,$$

where equality holds if and only if the triangle is equilateral [3], [4], we get:

$$2s - 4R - r \leqslant (6\sqrt{3} - 9)r.$$

Hence:

THEOREM 5. *If a triangle has an isoperimetric point, the inequality*

$$\sigma/s \geqslant \frac{1}{9}(3 + 2\sqrt{3}) \tag{10.1}$$

*holds with equality if and only if the triangle is equilateral.*

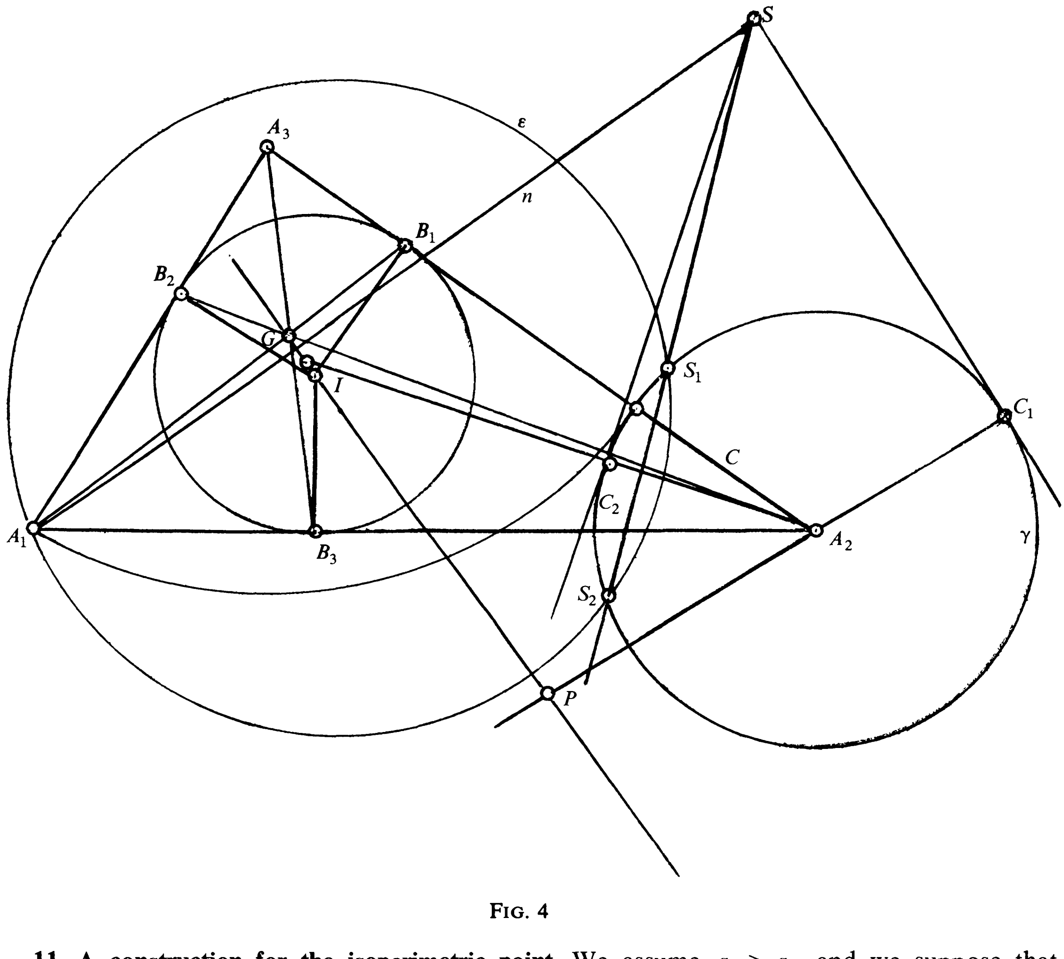



FIG. 4

**11. A construction for the isoperimetric point.** We assume $a_1 > a_2$ and we suppose that $\Delta A_1A_2A_3$ has an isoperimetric point $P$. It lies on $GI$ such that

$$\overline{PA_1} - \overline{PA_2} = \overline{A_3A_2} - \overline{A_3A_1}.$$

According to our assumption, this means that $P$ is the center of a circle $\delta$ passing through $A_1$ and having internal contact with the circle $\gamma$ with center $A_2$ and radius $c := \overline{A_3A_2} - \overline{A_3A_1}$. The set of all circles with centers on $GI$ and passing through $A_1$ is obviously a pencil with the perpendicular $n$ from $A_1$ onto $GI$ as its radical axis (Fig. 4). Let $\varepsilon$ be a member of this pencil meeting $\gamma$ at $S_1$

and $S_2$. The radical axis of $\varepsilon$ and $\gamma$ is $S_1S_2$; that of $\varepsilon$ and the circle $\delta$ we are looking for is $n$, whereas that of $\delta$ and $\gamma$ is a common tangent to these circles. Since these three radical axes are concurrent, this common tangent passes through the intersection $S$ of $n$ and $S_1S_2$. The points of contact $C_1$ and $C_2$ of $\gamma$ with its tangents through $S$ coincide with the points of contact of $\gamma$ with the two members of the pencil which are tangent to $\gamma$. At Fig. 4 the point $C_2$ corresponds to the member with external contact and $A_2C_2$ meets $GI$ therefore at the point which in the preceding section has been denoted by $Q$. The intersection of $A_2C_1$ and $GI$ is the isoperimetric point $P$ of the triangle.

**12. The geometric point of view.** Let $\Gamma_0$ be a circle concentric with $\gamma_0$ with radius exceeding $r$. The six intersections of $\Gamma_0$ with the side lines of $\Delta A_1A_2A_3$ are denoted by $C_{ij}$ in such a way that the segments $\overline{C_{12}C_{13}}$, $\overline{C_{23}C_{21}}$ and $\overline{C_{31}C_{32}}$ are of the same length $2p$ and have $B_1$, $B_2$ and $B_3$ as their midpoints, respectively, and that, moreover, $C_{ik}C_{jk} \| B_iB_j$ for any permutation of the indices (Fig. 5).

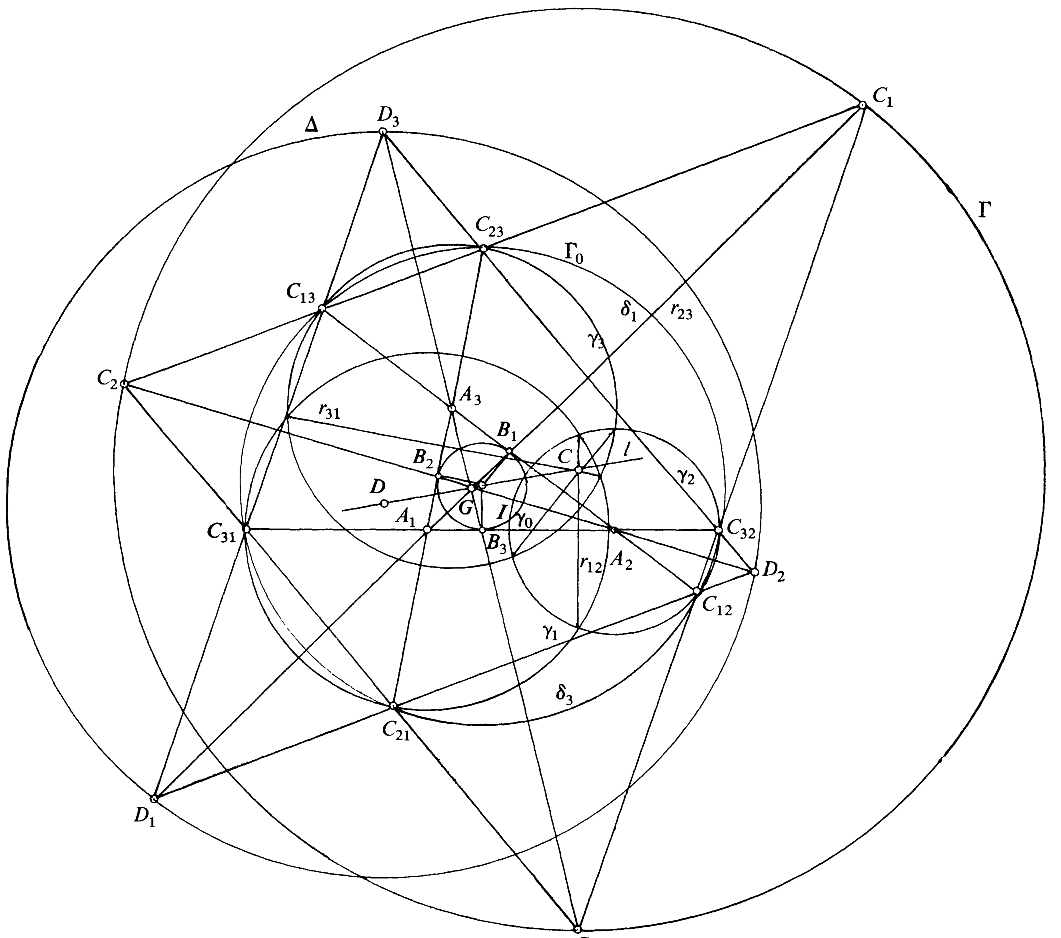



FIG. 5

Let $C_1, C_2, C_3$ and $D_1, D_2, D_3$ be in this order the intersections

$$C_{12}C_{32} \cap C_{13}C_{23}, C_{23}C_{13} \cap C_{21}C_{31}, C_{31}C_{21} \cap C_{32}C_{12}$$

and

$$C_{12}C_{21} \cap C_{13}C_{31}, C_{23}C_{32} \cap C_{21}C_{12}, C_{31}C_{13} \cap C_{32}C_{23}.$$

Then $\Delta C_1C_2C_3$ and $\Delta D_1D_2D_3$ are homothetic with $\Delta B_1B_2B_3$. The triangles $D_1C_{21}C_{31}$ and $B_1B_2B_3$ are homothetic too. Since $C_{21}B_2$ and $C_{31}B_3$ intersect at $A_1$, the line $D_1B_1$ passes through $A_1$. Moreover, $\Delta C_1C_{23}C_{32}$ is homothetic with $\Delta B_1B_2B_3$. The lines $C_{23}B_2$ and $C_{32}B_3$ have $A_1$ in common; hence $B_1C_1$ passes through $A_1$. It follows from these facts that $A_1$ and $B_1$ are on $C_1D_1$. Likewise $A_2$ and $B_2$ are on $C_2D_2$ and $A_3$ and $B_3$ on $C_3D_3$. Since $\Delta C_1C_2C_3$ and $\Delta D_1D_2D_3$ are homothetic, $C_1D_1$, $C_2D_2$ and $C_3D_3$ are concurrent. It follows that $A_1B_1$, $A_2B_2$ and $A_3B_3$ are concurrent. Their common point $G$ is known as the Gergonne point of $\Delta A_1A_2A_3$. Obviously $G$ is the center of homothecy for each of the 3 pairs of triangles taken from the triplet $\{\Delta B_1B_2B_3, \Delta C_1C_2C_3, \Delta D_1D_2D_3\}$.

We observe that $A_iC_{ji} = A_iC_{ki} = |p - (s - a_i)|$. The circle with center $A_i$ passing through $C_{ji}$ and $C_{ki}$ will be denoted by $\gamma_i$ ($i = 1,2,3$). The radical axes of the pairs $\{\Gamma, \gamma_2\}$ and $\{\Gamma, \gamma_3\}$ meet at $C_1$. Hence $C_1$ is on the radical axis of $\gamma_2$ and $\gamma_3$, this line being therefore the perpendicular from $C_1$ onto $l_1$. This means that the radical axis $r_{23}$ of $\gamma_2$ and $\gamma_3$ and the line $IB_1$ are corresponding lines under the homothecy between $\Delta C_1C_2C_3$ and $\Delta B_1B_2B_3$. The same holds for $r_{31}$ and $IB_2$, $r_{12}$ and $IB_3$. The radical center of $\gamma_1$, $\gamma_2$ and $\gamma_3$ (i.e., the common point $C$ of $r_{23}$, $r_{31}$ and $r_{12}$) therefore corresponds to $I$. So we have

THEOREM 6. *The radical center $C$ of the triplet $\{\gamma_1, \gamma_2, \gamma_3\}$ coincides with the center of the circumcircle $\Gamma$ of $C_1C_2C_3$ and lies on $GI$.*

If $\delta_i$ is the circle with center $A_i$ passing through $C_{jk}$ and $C_{kj}$, we can prove in a similar way

THEOREM 7. *The radical center $D$ of the triplet $\{\delta_1, \delta_2, \delta_3\}$ is the center of the circumcircle $\Delta$ of $\Delta D_1D_2D_3$ and lies on $GI$.*

**13. The powers of $C$ and $D$ with respect to the corresponding triplets.** The dilatation factors to obtain the triangles $A_1C_{21}C_{31}, C_{13}C_2C_{31}$ and $C_{12}C_{21}C_3$ from $\Delta B_1B_2B_3$ are in this order:

$$\frac{p-(s-a_1)}{s-a_1} = \frac{p}{s-a_1} - 1, \frac{p+s-a_2}{s-a_2} = \frac{p}{s-a_2} + 1, \frac{p+s-a_3}{s-a_3} = \frac{p}{s-a_3} + 1.$$

$\Delta C_1C_2C_3$ is therefore obtained from $\Delta B_1B_2B_3$ by the factor

$$\frac{p}{s-a_1} + \frac{p}{s-a_2} + \frac{p}{s-a_3} + 1 = \frac{p(r_1+r_2+r_3)}{F} + 1 = \frac{p(4R+r)}{F} + 1.$$

The radius $R_c$ of $\Gamma$ is therefore

$$R_c = \frac{p(4R+r)}{s} + r. \tag{13.1}$$

Using $\overline{B_2B_3^2} = 2(s-a_1)s_1$, we obtain furthermore

$$\overline{C_{31}C_3} = \overline{C_{31}C_{21}} + \overline{C_{21}C_3} = \left(\frac{p}{s-a_1} + \frac{p}{s-a_3}\right)\overline{B_2B_3} = \frac{ps}{2Rc_1c_3^2} \tag{13.2}$$

and

$$\overline{C_3C_{21}} = \left(\frac{p}{s-a_3} + 1\right)\overline{B_2B_3} = 2rc_1\left(\frac{p}{s-a_3} + 1\right).$$

The power of $C_3$ with respect to $\gamma_1$ is therefore

$$\overline{C_3A_1^2} - \rho_1^2 = \overline{C_3C_{31}} \times \overline{C_3C_{21}} = \frac{prs}{Rc_3^2}\left(\frac{p}{s-a_3} + 1\right), \tag{13.3}$$

where $\rho_1$ stands for the radius of $\gamma_1$.

Observing that $\angle C_{31}C_3C = \frac{1}{2}\alpha_1$ and applying the law of cosines in $\Delta CC_3A_1$, we obtain for the power $m_c$ of $C$ with respect to $\gamma_1$, in view of (13.2) and (13.3),

$$\begin{aligned} m_c &= \overline{CA}_1^2 - \rho_1^2 = R_c^2 + \overline{C_3A}_1^2 - 2R_c \cdot \overline{CA_1}\cos\angle A_1C_3C - \rho_1^2 \\ &= R_c^2 - 2R_c \cdot \overline{CA_1}\cos\angle A_1C_3C + \frac{prs}{Rc_3^2}\left(\frac{p}{s-a_3} + 1\right) \\ &= R_c^2 - \frac{ps}{Rc_3^2}R_c + \frac{prs}{Rc_3^2}\left(\frac{p}{s-a_3} + 1\right). \end{aligned}$$

Using (13.1), we get $m_c = Ap^2 + 2Bp + C$ with

$$A = \frac{(4R+r)^2}{s^2} - \frac{4R+r}{Rc_3^2(s-a_3)} = \frac{(4R+r)^2 - 4s^2}{s^2}; \quad B = \frac{2r(4R+r)}{s}; C = r^2.$$

It follows that

$$s^2 m_c = \{(4R + r + 2s)p + rs\}\{(4R + r - 2s)p + rs\}. \tag{13.4}$$

This result implies that there exists a value of $p$ such that $m_c$ vanishes if and only if $2s > 4R + r$, namely:

$$\sigma = \frac{rs}{2s - 4R - r}.$$

If $p = \sigma$, the circles $\gamma_1, \gamma_2, \gamma_3$ have one point in common: the isoperimetric point $P$ of the triangle. In this case we find from (13.1) for the dilatation factor from $\Delta B_1B_2B_3$ into $\Delta C_1C_2C_3$ the value $2s/(2s - 4R - r)$. *The radius of the circle* $\Gamma$ *is therefore in this special case equal to* $2\sigma$ *and its center coincides with* $P$.

The dilatation factor to transform $\Delta B_1B_2B_3$ into $\Delta D_1D_2D_3$ is $\left|\frac{p(4R+r)}{F} - 1\right|$ and the radius $R_D$ of $\Delta$ is therefore given by

$$R_D = \left|\frac{p(4R+r)}{s} - r\right|. \tag{13.5}$$

For the power of $D$ with respect to the triplet $\{\delta_1, \delta_2, \delta_3\}$ one has

$$s^2 m_D = \{(4R + r + 2s)p - rs\}\{(4R + r - 2s)p - rs\}. \tag{13.6}$$

It follows from the above that

$$\overline{GC} = \left\{\frac{p(4R+r)}{F} + 1\right\}\overline{GI} \quad \text{and} \quad \overline{GD} = \left\{\frac{p(4R+r)}{F} - 1\right\}\overline{GI}.$$

As a consequence, *the centers of* $\Gamma$ *and* $\Delta$ *lie symmetric with respect to* $I$.

### 14. Additional remarks.

1. We have for the point $Q$:

$$q_1 - q_2 = a_2 - a_1, \qquad q_2 - q_3 = a_3 - a_1, \quad \text{and} \quad q_3 - q_1 = a_1 - a_2.$$

This is equivalent to

$$q_2 + q_3 - a_1 = q_3 + q_1 - a_2 = q_1 + q_2 - a_3. \tag{14.1}$$

The shortest way from a point $A$ to a point $B$ is along the straight line $AB$. If $X$ is a point not between $A$ and $B$, we make a detour of magnitude $\overline{AX} + \overline{XB} - \overline{AB}$ if we walk from $A$ to $B$ via $X$. The equalities (14.1) are obviously the expression for the following property of the point $Q$: the

detour $\overline{QA_i} + \overline{QA_j} - \overline{A_iA_j}$ in travelling from $A_i$ to $A_j$ via $Q$ is the same for all pairs $\{i, j\}$ taken from $\{1,2,3\}$. Therefore $Q$ could be named a *point of equal detour* for $\Delta A_1A_2A_3$. Denoting this detour by $2\tau$, we find

$$3\tau = q_1 + q_2 + q_3 - s.$$

It is easily found (see §7 and §9) that

$$q_1 + q_2 + q_3 = \frac{(4R + 4r + 2s)s}{4R + r + 2s}$$

and therefore

$$\tau = \frac{rs}{4R + r + 2s}. \tag{14.2}$$

It can be shown that $4R + r + 2s \geqslant (6\sqrt{3} + 9)r$ and therefore

$$\tau/s \leqslant \frac{1}{9}(2\sqrt{3} - 3)$$

with equality if and only if the triangle is equilateral.

*Contrary to the isoperimetric point, a point of equal detour exists for any triangle.*

2. We see from (13.6) that $m_D = 0$ if $p = \tau$. Therefore by taking $p = \tau$ the triplet $\{\delta_1, \delta_2, \delta_3\}$ has a point in common, namely, the point $Q$ of equal detour. It is the center of the corresponding circle $\Delta$ with radius $2\tau$. Furthermore (13.6) shows that $Q$ is the only point of equal detour if the triangle has an isoperimetric point, because in this case $(4R + r - 2s)p - rs < 0$ for all $p \geqslant 0$.

*If, however, no isoperimetric point exists, that is, if* $2s < 4R + r$ or—*what is the same—the largest angle exceeds* $2\arcsin 4/5$, *there are two points of equal detour.* They are: $Q$ and the point $P^*$ corresponding to the value

$$\tau^* = \frac{rs}{4R + r - 2s}$$

of $p$, this value now being apart from $\tau$ a zero of the second member of (13.6). The corresponding triplet $\{\delta_1, \delta_2, \delta_3\}$ has $P^*$ as the common point, lying on the branch of $h_k$ passing through $A_k$ $(k = 1,2,3)$. The radius of $\Delta$ is in this case $2\tau^*$.

**Acknowledgement.** The author would like to express his sincere thanks to Dr. Karel Post for the encouragement to prepare this paper, to Mrs. E. W. van Thiel for her skillful typing, and to the scientific publication office of the Eindhoven University of Technology (Department of Mathematics) for the final preparation of the manuscript.

**References**

1. Howard Eves, A Survey of Geometry, Allyn and Bacon, Boston, 1965 (see esp. Problem 10.3-6).
2. Nathan Altshiller-Court, College Geometry, 2nd ed., Barnes and Noble, New York, 1952, pp. 160–162.
3. W. J. Blundon, Problem E 1935, this MONTHLY, 73 (1966) 1122.
4. O. Bottema, et al., Geometric Inequalities, Groningen, 1969, p. 49, inequality 5.4 (without proof).

---

## 154. MISCELLANEA

We are so accustomed to treating continuous processes as limits of discrete processes that we may tend to forget that this was once a startlingly original idea, put forth as a hypothesis. It is called Boscovich's Hypothesis. [1758]

—Richard Bellman, *Adaptive Control Processes: A Guided Tour*,
Princeton University Press, 1961, p. 80.

One of the members of the first husband and wife partnership in the history of mathematics. (See p. 578.)

(Notice that no terms $\binom{\beta}{k}$ enter into the coefficient of $x^n$.) ■

This theorem gives some insight into the $p$-adic expansion. It follows easily from this theorem that the $p$-adic expansion for $\alpha$ is

$$\alpha = \sum_{n=0}^{\infty}\left(\binom{\alpha}{p^n}\operatorname{mod} p\right)p^n.$$

From this expression one can obtain formulas for the coefficients of the $p$-adic expansion for $\alpha+\beta$ in terms of the coefficients of $\alpha$ and $\beta$. If $\alpha = \sum_{i=0}^{\infty} a_i p^i$, $\beta = \sum_{i=0}^{\infty} b_i p^i$, then $\alpha + \beta = \sum_{i=0}^{\infty} c_i p^i$ where

$$c_n = a_n + b_n + \sum_{s=0}^{n-1}\binom{a_s+b_s}{p}\sum_{i=s+1}^{n-1}\binom{a_i+b_i}{p-1}\operatorname{mod} p.$$

The details for the proof of this formula are left to the interested reader.

COROLLARY 1. *If $n = a_0 + a_1 p + \cdots + a_s p^s$ is a natural number, then*

$$\binom{2n}{n} = (-1)^n 2^{2n}\prod_{i=0}^{s}\binom{\frac{1}{2}(p-1)}{a_i}\operatorname{mod} p.$$

*Proof.* From the uniqueness of $p$-adic expansions and the geometric series for $1/(1-p)$, it follows that

$$-\frac{1}{2} = \frac{p-1}{2}(1 + p + p^2 + \cdots)$$

is the $p$-adic expansion of $-\frac{1}{2}$. Evaluating $\binom{-\frac{1}{2}}{n}$ by means of the theorem and combining it with the formula

$$\binom{-\frac{1}{2}}{n} = (-1)^n 2^{-2n}\binom{2n}{n}$$

we obtain the stated result. ■

According to Euler's Theorem, for an odd prime $p$,

$$\left(\frac{2}{p}\right) = 2^{\frac{1}{2}(p-1)}\operatorname{mod} p.$$

Using the next corollary we can obtain the evaluation of $\left(\frac{2}{p}\right)$.

COROLLARY 2. For an odd prime $p$, $2^{\frac{1}{2}(p-1)} = (-1)^{(p^2-1)/8}\operatorname{mod}$ p.

*Proof.* If $p = 1 \operatorname{mod} 4$, let $n = (p-1)/4$ and apply the previous corollary. If $p = -1 \operatorname{mod} 4$, let $n = (p+1)/4$ and apply the previous corollary. ■

## Reference

1. N. Koblitz, *P*-adic Numbers, *P*-adic Analysis and Zeta-Functions. Springer-Verlag, 1977.

---

## ANSWER TO PHOTO ON PAGE 559

The picture shows Grace Chisholm Young (1868–1944) who with her husband William Henry Young (1862–1943) produced fundamental research in general topology, measure theory, and functional analysis—and also six children, who include Professor L. C. Young of the University of Wisconsin. The picture was taken before her marriage, in fact probably in the mid 1890s when she took her doctorate at Göttingen, the first in the new program.

# THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

L. F. KLOSINSKI
*Department of Mathematics, University of Santa Clara, CA 95053*

G. L. ALEXANDERSON
*Department of Mathematics, University of Santa Clara, CA 95053*

L. C. LARSON
*Department of Mathematics, St. Olaf College, Northfield, MN 55057*

The following results of the forty-fifth William Lowell Putnam Mathematical Competition, held on December 1, 1984, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, five thousand dollars, was awarded to both the Department of Mathematics of the **University of California**, Davis, California, and the Department of Mathematics of **Washington University**, St. Louis, Missouri. The teams from these institutions tied. The members of the team from the University of California, Davis, were: John B. Boyland, Robert J. Filippini, and Michael P. Quinn. The members of the team from Washington University, St. Louis, were: William H. Paulsen, Richard A. Stong, Dougin A. Walker. Each member of these teams was awarded a prize of two hundred fifty dollars.

The third prize, one thousand five hundred dollars, was awarded to the Department of Mathematics of **Harvard University**, Cambridge, Massachusetts. The members of its team were: Benji N. Fisher, Howard M. Pollack, and John M. Sullivan; each was awarded a prize of one hundred fifty dollars.

The fourth prize, one thousand dollars, was awarded to the Department of Mathematics of **Princeton University**, Princeton, New Jersey. The members of the team were: Douglas R. Davidson, Gregg N. Patruno, and James C. Yeh; each was awarded a prize of one hundred dollars.

The fifth prize, five hundred dollars, was awarded to the Department of Mathematics of **Yale University**, New Haven, Connecticut. The members of its team were: Thomas O. Andrews, Nathaniel E. Glasser, and Stephen E. Mark; each was awarded a prize of fifty dollars.

The five highest-ranking individual contestants, in alphabetical order, were **Noam D. Elkies**, Columbia University; **Benji N. Fisher**, Harvard University; **Daniel W. Johnson**, Rose-Hulman Institute of Technology; **Michael Reid**, Harvard University; and **Richard A. Stong**, Washington University, St. Louis. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of five hundred dollars by the Putnam Prize Fund.

The next five highest ranking individuals, in alphabetical order, were *Leland F. Brown*, California Institute of Technology; *Douglas R. Davidson*, Princeton University; *Zachary M. Franco*, Harvard University; *John M. Sullivan*, Harvard University; and *David S. Yuen*, University of Chicago. Each of these students was awarded a prize of two hundred fifty dollars.

The following teams, named in alphabetical order, received honorable mention: *California Institute of Technology*, with team members Everett W. Howe, Jung C. Im, and Eric H. Kawamoto; *University of California*, Berkeley, with team members Michael J. McGrath, Jonathan E. Shapiro, and Christopher S. Welty; *University of Chicago*, with team members Eric K. Lossin, Keith A. Ramsay, and David S. Yuen; *Colorado State University*, with team members Curtis D. Bennett, Jorg A. Brown, and Mark S. Vincent; and *Rose-Hulman Institute of Technology*, with team members Todd G. Fine, Erich J. Friedman, and Daniel W. Johnson.

Honorable mention was achieved by the following thirty-five individuals, named in alphabeti-

cal order: *Michael A. Abramson*, Princeton University; *Wayne E. Aitken*, Brigham Young University; *Thomas O. Andrews*, Yale University; *Mark E. Banilower*, University of Pennsylvania; *Arthur B. Baragar*, University of Alberta; *William J. Bruno*, Massachusetts Institute of Technology; *François Destrempes*, Université de Montreal; *Art Duval*, California Institute of Technology; *Duane J. Einfeld*, Dordt College; *Robert J. Filippini*, University of California, Davis; *Leonid Fridman*, Harvard University; *Nathaniel E. Glasser*, Yale University; *Michelangelo Grigni*, Duke University; *George E. Homsy*, University of California, Berkeley; *Liaw Huang*, Rutgers University, New Brunswick; *Douglas S. Jungreis*, Harvard University; *Joe J. Kilian*, Massachusetts Institute of Technology; *Tsz Mei Ko*, The Cooper Union; *Rama R. Kocherlakota*, Princeton University; *Chun-Nip Lee*, Massachusetts Institute of Technology; *Tak Kwan Lee*, University of California, San Diego; *Daniel E. Loeb*, California Institute of Technology; *Michael J. McGrath*, University of California, Berkeley; *Mark E. Meyer*, Indiana University; *Lee A. Newberg*, Massachusetts Institute of Technology; *Jeremy D. Primer*, Princeton University; *Michael P. Quinn*, University of California, Davis; *Keith A. Ramsay*, University of Chicago; *Daniel N. Ropp*, Washington University, St. Louis; *Alistair M. Rucklidge*, University of Toronto; *David S. Salopek*, University of Alberta; *Kenneth W. Shirriff*, University of Waterloo; *Charles S. A. Timar*, University of Waterloo; *Luis G. Valdez-Sanchez*, University of Texas, El Paso; and *Dougin A. Walker*, Washington University, St. Louis.

The other individuals who achieved ranks among the top 102, in alphabetical order of their schools, were: Brigham Young University, *Christopher P. Grant*; California Institute of Technology, *Eric K. Babson*, *William D. Banks*, *Kent J. Cantwell*, *Karen L. Condie*, *William D. Cutrell*, *Jung C. Im*, *James T. Liu*, *Julian West*, *Tad P. White*; University of California, Davis, *John B. Boyland*; University of California, Santa Barbara, *Emerson S. Fang*; Calvin College, *Randy V. Gritter*; Carnegie-Mellon University, *Steven G. DesJardins*, *Jamshid Mahdavi*; University of Chicago, *Geoffrey R. Harris*, *Rainer Hollerbach*, *Thomas R. Lippincott*, *Susan Tolman*; Colorado State University, *Curtis D. Bennett*, *Jorg A. Brown*; Concordia University (G. S. W.), *Chinh Mai*; Cornell University, *Denise E. Freed*; Harvard University, *Glenn D. Ellison*; Haverford College, *Kian-Tat Lim*; University of Illinois, Urbana, *Mark A. Thompson*; University of Kansas, *Glenn G. Chappell*, *Ryan D. Moats*; Université Laval, *Mario M. B. Bergeron*; University of Louisville, *Dung T. Nguyen*; Massachusetts Institute of Technology, *Avrim L. Blum*; Michigan State University, *Frank Sottile*; University of Michigan, Ann Arbor, *Steve Newman*; University of Minnesota, Minneapolis, *Jay A. Jorgenson*; University of Missouri, Rolla, *Ervan E. Darnell*; University of Nebraska, Lincoln, *Bartley E. Goddard*; North Dakota State University, *Jim B. Becker*; University of Notre Dame, *James R. Roche*; Princeton University, *Gregg N. Patruno*, *James C. Yeh*; Rice University, *Garrett T. Biehle*; Rutgers University, New Brunswick, *Scott E. Axelrod*; Simon Fraser University, *Stuart G. Cowan*; Swarthmore College, *John H. Palmieri*; University of Toronto, *Gary F. Baumgartner*, *William J. Rucklidge*; University of Victoria, *Philip H. Spencer*; University of Washington, *Charles N. Curtis*; Washington University, St. Louis, *William H. Paulsen*; University of Waterloo, *Yong Yao Du*, *Alexander T. Kachura*; Wichita State University, *Paul D. Sinclair*; Williams College, *Martin V. Hildebrand*; University of Wisconsin, Eau Claire, *Kenneth J. Dykema*; University of Wisconsin, Milwaukee, *Mark W. Hopkins*; and Yale University, *Ramzi R. Khuri*, *David R. Steinsaltz*.

There were 2149 individual contestants from 350 colleges and universities in Canada and the United States in the competition of December 1, 1984. Teams were entered by 264 institutions.

The Questions Committee for the forty-fifth competition consisted of Melvin Hochster (Chairman), Bruce Reznick, and Richard P. Stanley; they composed the problems listed below and were most prominent among those suggesting solutions.

## PROBLEMS

**Problem A-1**

Let $A$ be a solid $a \times b \times c$ rectangular brick in three dimensions, where $a, b, c > 0$. Let $B$ be the set of all

points which are a distance at most one from some point of $A$ (in particular, $B$ contains $A$). Express the volume of $B$ as a polynomial in $a$, $b$, and $c$.

**Problem A-2**

Express $\Sigma_{k=1}^{\infty}(6^k/(3^{k+1}-2^{k+1})(3^k-2^k))$ as a rational number.

**Problem A-3**

Let $n$ be a positive integer. Let $a, b, x$ be real numbers, with $a \neq b$, and let $M_n$ denote the $2n \times 2n$ matrix whose $(i, j)$ entry $m_{ij}$ is given by

$$m_{ij} = \begin{cases} x & \text{if } i = j, \\ a & \text{if } i \neq j \quad \text{and} \quad i + j \text{ is even,} \\ b & \text{if } i \neq j \quad \text{and} \quad i + j \text{ is odd.} \end{cases}$$

Thus, for example, $M_2 = \begin{pmatrix} x & b & a & b \\ b & x & b & a \\ a & b & x & b \\ b & a & b & x \end{pmatrix}$. Express $\lim_{x \to a} \det M_n/(x-a)^{2n-2}$ as a polynomial in $a$, $b$, and $n$, where $\det M_n$ denotes the determinant of $M_n$.

**Problem A-4**

A convex pentagon $P = ABCDE$, with vertices labeled consecutively, is inscribed in a circle of radius 1. Find the maximum area of $P$ subject to the condition that the chords $AC$ and $BD$ be perpendicular.

**Problem A-5**

Let $R$ be the region consisting of all triples $(x, y, z)$ of nonnegative real numbers satisfying $x + y + z \leqslant 1$. Let $w = 1 - x - y - z$. Express the value of the triple integral

$$\iiint_R x^1 y^9 z^8 w^4 \, dx \, dy \, dz$$

in the form $a!b!c!d!/n!$, where $a$, $b$, $c$, $d$, and $n$ are positive integers.

**Problem A-6.** Let $n$ be a positive integer, and let $f(n)$ denote the last nonzero digit in the decimal expansion of $n!$. For instance, $f(5) = 2$.

(a) Show that if $a_1, a_2, \ldots, a_k$ are *distinct* nonnegative integers, then $f(5^{a_1} + 5^{a_2} + \cdots + 5^{a_k})$ depends only on the sum $a_1 + a_2 + \cdots + a_k$.

(b) Assuming part (a), we can define

$$g(s) = f(5^{a_1} + 5^{a_2} + \cdots + 5^{a_k}),$$

where $s = a_1 + a_2 + \cdots + a_k$. Find the least positive integer $p$ for which

$$g(s) = g(s + p), \quad \text{for all } s \geqslant 1,$$

or else show that no such $p$ exists.

**Problem B-1**

Let $n$ be a positive integer, and define

$$f(n) = 1! + 2! + \cdots + n!.$$

Find polynomials $P(x)$ and $Q(x)$ such that

$$f(n+2) = P(n)f(n+1) + Q(n)f(n),$$

for all $n \geqslant 1$.

**Problem B-2.** Find the minimum value of

$$(u - v)^2 + \left(\sqrt{2 - u^2} - \frac{9}{v}\right)^2$$

for $0 < u < \sqrt{2}$ and $v > 0$.

**Problem B-3**

Prove or disprove the following statement: If $F$ is a finite set with two or more elements, then there exists a binary operation $*$ on $F$ such that for all $x, y, z$ in $F$,

(i) $x * z = y * z$ implies $x = y$ (right cancellation holds),

and

(ii) $x * (y * z) \neq (x * y) * z$ (*no* case of associativity holds).

**Problem B-4**

Find, with proof, all real-valued functions $y = g(x)$ defined and *continuous* on $[0, \infty)$, positive on $(0, \infty)$, such that for all $x > 0$ the $y$-coordinate of the centroid of the region

$$R_x = \{(s,t) | 0 \leqslant s \leqslant x, \quad 0 \leqslant t \leqslant g(s)\}$$

is the same as the average value of $g$ on $[0, x]$.

**Problem B-5**

For each nonnegative integer $k$, let $d(k)$ denote the number of 1's in the binary expansion of $k$ (for example, $d(0) = 0$ and $d(5) = 2$). Let $m$ be a positive integer. Express

$$\sum_{k=0}^{2^m-1} (-1)^{d(k)} k^m$$

in the form $(-1)^m a^{f(m)}(g(m))!$, where $a$ is an integer and $f$ and $g$ are polynomials.

**Problem B-6**

A sequence of convex polygons $\{P_n\}$, $n \geqslant 0$, is defined inductively as follows. $P_0$ is an equilateral triangle with sides of length 1. Once $P_n$ has been determined, its sides are trisected; the vertices of $P_{n+1}$ are the *interior* trisection points of the sides of $P_n$. Thus, $P_{n+1}$ is obtained by cutting corners off $P_n$, and $P_n$ has $3 \cdot 2^n$ sides. ($P_1$ is a regular hexagon with sides of length $1/3$.)

Express $\lim_{n\to\infty} \text{Area}(P_n)$ in the form $\sqrt{a}/b$, where $a$ and $b$ are positive integers.

## SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, $n_i$ for $10 \geqslant i \geqslant 0$ is the number of students among the top 205 contestants achieving $i$ points for the problem and $n_{-1}$ is the number of those not submitting solutions.

**A-1.** (146, 3, 33, 0, 0, 0, 0, 0, 21, 1, 1, 0)

The set $B$ can be partitioned into the following sets:

(i) A itself, of volume $abc$;

(ii) two $a \times b \times 1$ bricks, two $a \times c \times 1$ bricks, and two $b \times c \times 1$ bricks, of total volume $2ab + 2ac + 2bc$;

(iii) four quarter-cylinders of length $a$ and radius 1, four quarter-cylinders of length $b$ and radius 1, and four quarter-cylinders of length $c$ and radius 1, of total volume $(a + b + c)\pi$;

(iv) eight spherical sectors, each consisting of one-eighth of a sphere of radius 1, of total volume $4\pi/3$.

Hence the volume of $B$ is

$$abc + 2(ab + ac + bc) + \pi(a + b + c) + \frac{4\pi}{3}.$$

**A-2.** (73, 3, 12, 1, 0, 11, 3, 17, 2, 1, 16, 66)

Let $S(n)$ denote the $n$th partial sum of the given series. Then

$$S(n)=\sum_{k=1}^{n}\left[\frac{3^k}{3^k-2^k}-\frac{3^{k+1}}{3^{k+1}-2^{k+1}}\right]=3-\frac{3^{n+1}}{3^{n+1}-2^{n+1}},$$

and the series converges to $\lim_{n\to\infty}S(n)=2$.

**A-3.** (38, 1, 7, 0, 0, 0, 0, 1, 7, 0, 55, 96)

Let $N=M_n]_{x=a}$. $N$ has rank 2, so that 0 is an eigenvalue of multiplicity $2n-2$. Let $\mathbf{e}$ denote the $2n\times 1$ column vector of 1's. Notice that $N\mathbf{e}=n(a+b)\mathbf{e}$, and therefore $n(a+b)$ is an eigenvalue. The trace of $N$ is $2na$, and therefore the remaining eigenvalue is $2na-n(a+b)=n(a-b)$. [Note: This corresponds to the eigenvector $\mathbf{f}$, where $f_{i,1}=(-1)^{i+1}$, $i=1,\dots,2n$.]

The preceding analysis implies that the characteristic equation of $N$ is

$$\det(N-\lambda I)=\lambda^{2n-2}(\lambda-n(a+b))(\lambda-n(a-b)).$$

Let $\lambda=a-x$. Then

$$\det M_n=\det(N-(a-x)I)=(a-x)^{2n-2}(a-x-n(a+b))(a-x-n(a-b)).$$

It follows that

$$\lim_{x\to a}\frac{\det M_n}{(x-a)^{2n-2}}=\lim_{x\to a}(a-x-n(a+b))(a-x-n(a-b))=n^2(a^2-b^2).$$

**A-4.** (7, 3, 4, 3, 0, 0, 0, 6, 3, 4, 82, 93)

Let $\theta=\operatorname{Arc}AB$, $\alpha=\operatorname{Arc}DE$, and $\beta=\operatorname{Arc}EA$. Then $\operatorname{Arc}CD=\pi-\theta$ and $\operatorname{Arc}BC=\pi-\alpha-\beta$.

The area of $P$, in terms of the five triangles from the center of the circle is

$$\frac{1}{2}\sin\theta+\frac{1}{2}\sin(\pi-\theta)+\frac{1}{2}\sin\alpha+\frac{1}{2}\sin\beta+\frac{1}{2}\sin(\pi-\alpha-\beta).$$

This is maximized when $\theta=\pi/2$ and $\alpha=\beta=\pi/3$. Thus, the maximum area is

$$\frac{1}{2}\cdot 1+\frac{1}{2}\cdot 1+\frac{1}{2}\frac{\sqrt{3}}{2}+\frac{1}{2}\frac{\sqrt{3}}{2}+\frac{1}{2}\frac{\sqrt{3}}{2}=1+\frac{3}{4}\sqrt{3}.$$

**A-5.** (52, 4, 4, 1, 0, 1, 0, 0, 5, 3, 39, 96)

For $t>0$, let $R_t$ be the region consisting of all triples $(x,y,z)$ of nonnegative real numbers satisfying $x+y+z\leqslant t$. Let

$$I(t)=\iiint_{R_t}x^1y^9z^8(t-x-y-z)^4\,dx\,dy\,dz$$

and make the change of variables $x=tu$, $y=tv$, $z=tw$. We see that $I(t)=I(1)t^{25}$.

Let $J=\int_0^\infty I(t)e^{-t}\,dt$. Then

$$J=\int_0^\infty I(1)\,t^{25}e^{-t}\,dt=I(1)\,\Gamma(26)=I(1)25!.$$

It is also the case that

$$J=\int_{t=0}^\infty\iiint_{R_t}e^{-t}x^1y^9z^8(t-x-y-z)^4\,dx\,dy\,dz\,dt.$$

Let $s=t-x-y-z$. Then

$$J=\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty e^{-s}e^{-x}e^{-y}e^{-z}x^1y^9z^8s^4\,dx\,dy\,dz\,ds=\Gamma(2)\,\Gamma(10)\,\Gamma(9)\,\Gamma(5)=1!9!8!4!.$$

The integral we desire is $I(1)=J/25!=1!9!8!4!/25!$.

**A-6.** (3, 1, 0, 1, 1, 0, 0, 1, 2, 0, 23, 173)

(a) All congruences are modulo 10.

LEMMA. $f(5n) \equiv 2^n f(n)$.

*Proof.* We have

$$(*) \qquad (5n)! = 10^n n! \prod_{i=0}^{n-1} \frac{(5i+1)(5i+2)(5i+3)(5i+4)}{2}.$$

If $i$ is even, then

$$\frac{1}{2}(5i+1)(5i+2)(5i+3)(5i+4) \equiv \frac{1}{2}(1 \cdot 2 \cdot 3 \cdot 4) \equiv 2,$$

and if $i$ is odd, then

$$\frac{1}{2}(5i+1)(5i+2)(5i+3)(5i+4) \equiv \frac{1}{2}(6 \cdot 7 \cdot 8 \cdot 9) \equiv 2.$$

Thus the entire product above is congruent to $2^n$. From $(*)$ it is clear that the largest power of 10 dividing $(5n)!$ is the same as the largest power of 10 dividing $10^n n!$, and the proof follows.

We now show by induction on $5^{a_1} + \cdots + 5^{a_k}$ that

$$f(5^{a_1} + \cdots + 5^{a_k}) \equiv 2^{a_1 + \cdots + a_k}$$

(which depends only on $a_1 + \cdots + a_k$ as desired).
This is true for $5^{a_1} + \cdots + 5^{a_k} = 1$, since $f(5^0) \equiv 2^0 \equiv 1$.

CASE 1. All $a_i > 0$. By the lemma and induction,

$$\begin{aligned} f(5^{a_1} + \cdots + 5^{a_k}) &\equiv 2^{5^{a_1-1} + \cdots + 5^{a_k-1}} f(5^{a_1-1} + \cdots + 5^{a_k-1}) \\ &\equiv 2^k \cdot 2^{(a_1-1) + \cdots + (a_k-1)} \qquad (\text{since } 2^{5^i} \equiv 2 \text{ for } i \geqslant 0) \\ &\equiv 2^{a_1 + \cdots + a_k}. \end{aligned}$$

CASE 2. Some $a_i = 0$, say $a_1 = 0$. Now

$$(1+5m)! = (1+5m)(5m)!,$$

so $f(1+5m) \equiv (1+5m)f(5m)$. But $f(5m)$ is even for $m \geqslant 1$ since $(5m)!$ is divisible by a higher power of 2 than of 5. But

$$(1+5m) \cdot (2j) \equiv 2j,$$

so $f(1+5m) \equiv f(5m)$. Letting $m = 5^{a_2-1} + \cdots + 5^{a_k-1}$, the proof follows by induction.

(b) The least $p \geqslant 1$ for which $2^{s+p} \equiv 2^2$ for all $s \geqslant 1$ is $p = 4$.

**B-1.** (179, 9, 6, 0, 0, 0, 0, 1, 0, 0, 4, 6)

We have

$$f(n+2) - f(n+1) = (n+2)! = (n+2)(n+1)! = (n+2)[f(n+1) - f(n)].$$

It follows that we can take $P(x) = x + 3$ and $Q(x) = -x - 2$.

**B-2.** (62, 13, 19, 6, 0, 0, 0, 0, 15, 15, 34, 41)

The problem asks for the minimum distance between the quarter of the circle $x^2 + y^2 = 2$ in the open first quadrant and the half of the hyperbola $xy = 9$ in that quadrant. Since the tangents to the respective curves at $(1,1)$ and $(3,3)$ separate the curves and are both perpendicular to $x = y$, the minimum distance is 8.

**B-3.** (93, 16, 1, 1, 0, 0, 0, 1, 2, 5, 33, 53)

The statement is true. Let $\phi$ be any bijection on $F$ with no fixed points, and set $x * y = \phi(x)$.

**B-4.** (2, 11, 2, 8, 0, 1, 0, 25, 4, 7, 48, 97)

Such a function must satisfy

$$\frac{\int_0^x \frac{1}{2} g^2(t)\, dt}{\int_0^x g(t)\, dt} = \frac{1}{x}\int_0^x g(t)\, dt,$$

or equivalently,

$$\int_0^x \frac{1}{2} g^2(t)\, dt = \frac{1}{x}\left[\int_0^x g(t)\, dt\right]^2.$$

Let $z(x) = \int_0^x g(t)\, dt$. Then $z'(x) = g(x)$ and we have

$$\int_0^x \frac{1}{2}(z')^2\, dt = \frac{z^2}{x}, \qquad x > 0.$$

Differentiating, we have

$$\frac{1}{2}(z')^2 = \frac{x \cdot 2zz' - z^2}{x^2}, \qquad x > 0,$$

$$x^2(z')^2 - 4xzz' + 2z^2 = 0, \qquad x > 0,$$

$$(xz' - r_1 z)(xz' - r_2 z) = 0, \qquad x > 0,$$

where $r_1 = 2 + \sqrt{2}$ and $r_2 = 2 - \sqrt{2}$.

Now $x$, $z'$, and $z$ are continuous and $z > 0$, so the last equation implies t $r = r_1$ or $r = r_2$. Separating variables, we have $z'/z = r/x$ and it follows tl

$$\ln z = r \ln x + C_0,$$

or equivalently, $z = C_1 x^r$, $C_1 > 0$. Differentiating, we have $z' = g(x) = Cx^{r-1}$, $C > 0$. But $g$ is continuous on $[0, \infty)$ and therefore we cannot have $r = r_2$ (because $r_2 - 1 = 1 - \sqrt{2} < 0$). Thus

$$g(x) = Cx^{1+\sqrt{2}}, \quad C > 0,$$

and one can check that such $g(x)$ do satisfy all the conditions of the problem.

**B-5.** (5, 1, 2, 0, 0, 3, 0, 4, 9, 11, 17, 153)

Define

$$D(x) = (1-x)(1-x^2)(1-x^4)\cdots(1-x^{2^{n-1}}).$$

Since binary expansions are unique, each monomial $x^k$ $(0 \leqslant k \leqslant 2^n - 1)$ appears exactly once in the expansion of $D(x)$, with coefficient $(-1)^{d(k)}$. That is,

$$D(x) = \sum_{k=0}^{2^n-1} (-1)^{d(k)} x^k.$$

Applying the operator $\left(x\frac{d}{dx}\right)$ to $D(x)$ $m$ times, we obtain

$$\left(x\frac{d}{dx}\right)^m D(x) = \sum_{k=0}^{2^n-1} (-1)^{d(k)} k^m x^k,$$

so that

$$\left(x\frac{d}{dx}\right)^m D(x)\bigg]_{x=1} = \sum_{k=0}^{2^n-1}(-1)^{d(k)}k^m.$$

Define $F(x) = D(x+1)$, so that

$$\left(x\frac{d}{dx}\right)^m D(x)\bigg]_{x=1} = \left[(x+1)\frac{d}{dx}\right]^m F(x)\bigg]_{x=0}.$$

But

$$F(x) = \prod_{\alpha=1}^{m}\left[1-(x+1)^{2^{\alpha-1}}\right] = \prod_{\alpha=1}^{m}\left[-2^{\alpha-1}x + O(x^2)\right], \quad (x\to 0),$$

$$= (-1)^m 2^{m(m-1)/2}x^m + O(x^{m+1}),$$

and by observing that $[(x+1)d/dx]x^n = nx^n + nx^{n-1}$, we see that

$$\left[(x+1)\frac{d}{dx}\right]^m\left(Ax^m + O(x^{m+1})\right) = m!A + O(x).$$

So

$$\left[(x+1)\frac{d}{dx}\right]^m F(x)\bigg]_{x=0} = (-1)^m 2^{m(m-1)/2}m! + O(x)\big]_{x=0} = (-1)^m 2^{m(m-1)/2}m!.$$

**B-6.** (7, 3, 3, 0, 0, 0, 0, 0, 6, 1, 48, 137)

Suppose that $\vec{u}$ and $\vec{v}$ are consecutive edges in $P_n$. Then $\vec{u}/3$, $(\vec{u}+\vec{v})/3$, and $\vec{v}/3$ are consecutive edges in $P_{n+1}$. Further,

$$\frac{1}{2}\left\|\frac{\vec{u}}{3}\times\frac{\vec{v}}{3}\right\| = \frac{1}{18}\|\vec{u}\times\vec{v}\|$$

is removed at this corner in making $P_{n+1}$. But at the next step, the amount from these three consecutive edges is

$$\frac{1}{2}\left\|\frac{\vec{u}}{9}\times\frac{\vec{u}+\vec{v}}{9}\right\| + \frac{1}{2}\left\|\frac{\vec{u}+\vec{v}}{9}\times\frac{\vec{v}}{9}\right\| = \frac{1}{81}\|\vec{u}\times\vec{v}\|.$$

Thus, the amount removed in the $(k+1)$st snip is $2/9$ times the amount removed in the $k$th.

Note that one-third of the original area is removed at the first step. Thus, the amount removed altogether is

$$\frac{1}{3}\left[1+(2/9)+(2/9)^2+\cdots\right] = \frac{1}{3}\cdot\frac{9}{7} = \frac{3}{7}$$

of the original area. Since the original area is $\sqrt{3}/4$, we have

$$\lim_{n\to\infty}\text{Area } P_n = \frac{4}{7}\cdot\frac{\sqrt{3}}{4} = \frac{\sqrt{3}}{7}.$$

The curve in this problem has been studied extensively by Georges de Rham. (See "Un peu de mathématiques à propos d'une courbe plane," *Elem. Math.*, 2 (1947), 73–76, 89–97; "Sur une courbe plane," *J. Math. Pures Appl.*, 35 (1956), 25–42; and "Sur les courbes limites de polygones obtenus par trisection," *Enseign. Math.*, 5 (1959), 29–43.) Among de Rham's results are the following. The limiting curve is $C^1$ with zero curvature almost everywhere, but every subarc contains points where the curvature is infinite. Consequently, the curve is nowhere analytic. De Rham parametrizes pieces of the curve so that the tangent vector is intimately related to the Minkowski ?-function. If the construction is repeated, but with each edge divided in the ratio (1/4, 1/2, 1/4) rather than (1/3, 1/3, 1/3), then the resulting limit curve *is* analytic, consisting of piecewise parabolic arcs.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## HOW DOES A COMPLETE GRAPH SPLIT INTO BIPARTITE GRAPHS AND HOW ARE NEIGHBORLY CUBES ARRANGED?

JOSEPH ZAKS

*Department of Mathematics, University of Haifa, Haifa 31999, Israel*

The **complete graph** $K_n$ has $n$ vertices and every pair of vertices is joined by an edge. The **complete bipartite graph** $K_{n,m}$ has $n$ vertices of one type and $m$ vertices of another type, and it has $mn$ edges, joining every vertex of one type to every vertex of the other type.

We are interested in splitting the edge-set of $K_n$ into parts, so that each part will consist of the full edge-set of some complete bipartite subgraph of $K_n$. The equation

$$K_n = \sum x_{ij} K_{i,j} \tag{1}$$

will mean that $K_n$ is decomposed into $x_{ij}$ copies of complete bipartite subgraphs $K_{i,j}$, where $i, j \geqslant 1$, $i + j \leqslant n$ and $x_{ij}$ are nonnegative integers.

A theorem of Graham & Pollak [**1**] (for shorter proofs, see [**5**] and [**3**]) states that $K_n$ can be decomposed into not fewer than $n - 1$ complete bipartite subgraphs; this theorem was extended recently by Reznick, Tiwari & West [**4**]. We restate this result.

THEOREM 1. $K_n = \Sigma x_{ij} K_{i,j}$ *implies that*

$$\sum x_{ij} \geqslant n - 1. \tag{2}$$

By counting the total number of edges, we easily derive (see also [**6**])

THEOREM 2. $K_n = \Sigma x_{ij} K_{i,j}$ *implies that*

$$\sum ij x_{ij} = \binom{n}{2}. \tag{3}$$

We raise the following open question.

PROBLEM. Given $n$ and a set of nonnegative integers $\{x_{ij} | i, j \geqslant 1,\ i + j \leqslant n\}$, satisfying the equation (3) and the inequality (2), when is there a decomposition of $K_n$ of the form (1)?

The conditions (2) and (3) on $n$ and the $x_{ij}$ are far from being sufficient: If $K_{m,m}$ is imbedded in $K_{2m}$, the remaining graph is the disjoint union of two copies of $K_m$; therefore $K_{2m}$ has no decomposition of the form $K_{m,m} + K_{a,b} + \cdots$, where $a + b \geqslant m + 1$.

A more interesting example is the following.

CLAIM. $K_9 \neq 6K_{2,2} + 2K_{2,3}$ (remark that the right-hand side has at least as many components as required and as many edges as needed.).

*Proof*. All the vertices of $K_9$ and of $K_{2,2}$ have even valence (number of edges having that vertex as an end-vertex); if $K_9 = 6K_{2,2} + 2K_{2,3}$, then the pairs of vertices of the two copies of $K_{2,3}$, which are the only vertices of odd valence in the decomposition, must coincide; therefore

$2K_{2,3}$ is actually equal to a $K_{2,6}$. Hence $K_9 = 6K_{2,2} + K_{2,6}$, which contradicts Theorem 1.

More generally,

$$K_{16m+9} \neq (16m + 6)K_{2,4m+2} + 2K_{2,6m+3} \quad \text{for all } m \geqslant 0.$$

Let us consider next the $d$-cube

$$Q = \{(x_1, \dots, x_d) \mid -1 \leqslant x_i \leqslant 1 \quad \text{for all } i, 1 \leqslant i \leqslant d\}$$

in the Euclidean $d$-space $E^d$. A **translation** of the $d$-cube $Q$ is a set of the form $v + Q = \{v + y \mid y \in Q\}$. Let $F = \{Q_1, Q_2, \dots\}$ be a collection of translations of the $d$-cube $Q$; $F$ is called **neighborly** (see **[6]** and the many references there) if every two members of $F$ intersect in a $(d-1)$-dimensional set; i.e., if every two members of $F$ are non-overlapping, yet they share a $(d-1)$-dimensional part of a facet of each.

There is a connexion between the decompositions of $K_n$ into complete bipartite graphs and neighborly families of translations of the $d$-cube in $E^d$.

THEOREM 3. *There is a* 1-1 *correspondence between the decompositions of $K_n$ into d complete bipartite graphs and the combinatorial type of neighborly families of n translations of the d-cube in $E^d$.*

*Proof.* If $Q_1 = (y_1, \dots, y_d) + Q$ and $Q_2 = (z_1, \dots, z_d) + Q$ are two translations of $Q$ which are neighborly, then $Q_1 \cap Q_2$ is a $(d-1)$-dimensional set which is convex; therefore $Q_1 \cap Q_2$ lies in a hyperplane which is parallel to one of the major hyperplanes in $E^d$, given by the equation $x_i = 0$, for some $i, 1 \leqslant i \leqslant d$; the following then holds

$$(4) \qquad |z_i - y_i| = 2, \quad \text{and} \quad |z_j - y_j| < 2 \quad \text{for all } j \neq i, \quad 1 \leqslant j \leqslant d.$$

In fact, condition (4) for some $i$ is a necessary and sufficient condition for two translations $(y_1, \dots, y_d) + Q$ and $(z_1, \dots, z_d) + Q$ to be neighborly.

Suppose $F = \{Q_1, \dots, Q_n\}$ is a neighborly family of translations of the $d$-cube $Q$; $Q_i = (y_{i1}, \dots, y_{id}) + Q$ for all $i, 1 \leqslant i \leqslant n$. Consider the $n \times d$ matrix $Y = (y_{ij})$, formed by the centres of the cubes. By the neighborliness of $F$ and condition (4) we get

(5) For every two rows $i$ and $k$ of $Y$, there is a unique column $j$ of $Y$ such that $|y_{ij} - y_{kj}| = 2$ and $|y_{it} - y_{kt}| < 2$ for all $t \neq j, 1 \leqslant t \leqslant d$.

Construct a graph $G$, which might look as if it has multiple edges, as follows: $G$ has the numbers $1, 2, \dots, n$ for its $n$ vertices, in 1-1 correspondence with the rows of $Y$; two vertices $i$ and $k$ of $G$ are joined by an edge whenever there is a column $j$ in $Y$ for which $|y_{ij} - y_{kj}| = 2$. It follows from (5) that there are no multiple edges in $G$ (because of the uniqueness of $j$), and in fact $G$ is the complete graph $K_n$ (because every $i$ and $k$ are connected by an edge).

Consider an arbitrary column of $Y$, say the $j$th column, and the edges which it alone contributes to $K_n$. If $\max_{i,k}|y_{ij} - y_{kj}| < 2$, then the $j$th column does not contribute any edges, and the corresponding complete bipartite subgraph is taken to be empty. If

$$\max_{i,k}|y_{ij} - y_{kj}| = 2,$$

we let

$$A_j = \{t \mid y_{tj} = \min_i y_{ij}\} \quad \text{and} \quad B_j = \{t \mid y_{tj} = \max_i y_{ij}\}.$$

The $j$th column contributes $|A_j| \cdot |B_j|$ edges to $K_n$, one edge for every pair of indices $(a, b)$ where $a \in A_j$ and $b \in B_j$; in fact, all of these edges form a complete bipartite subgraph of $K_n$, which we denote by $K_{A_j, B_j}$.

It follows that $K_n = \sum_{j=1}^{d} K_{A_j, B_j}$ is the decomposition of $K_n$ which is associated with $F$.

Suppose, on the other hand, that a decomposition of $K_n$ into $d$ complete bipartite subgraphs is given; to simplify the notation, let the vertex set of $K_n$ be $\{1, 2, \dots, n\}$ and let $K_n = \sum_{j=1}^{d} K_{A_j, B_j}$

where $A_j, B_j \subset \{1,2,\dots,n\}$ and $A_j \cap B_j = \varnothing$. Form the $n \times d$ matrix $Z = (z_{ij})$ by defining

$$z_{ij} = \begin{cases} 0 & \text{if } i \in A_j, \\ 2 & \text{if } i \in B_j, \\ 1 & \text{if } i \notin A_j \cup B_j. \end{cases}$$

The corresponding neighborly family of $n$ translations of the $d$-cube $Q$ in $E^d$ is obtained by using the rows of $Z$ for centres; i.e., the $i$th cube in the family is given by $(z_{i1}, z_{i2}, \dots, z_{id}) + Q$, $1 \leqslant i \leqslant n$. The neighborliness follows easily from (4).

This completes the proof of Theorem 3.

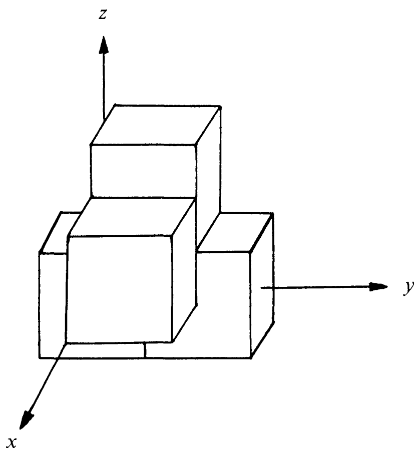



FIG. 1

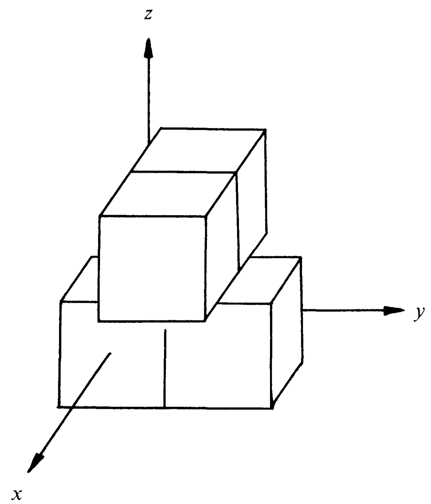



FIG. 2

To illustrate the correspondence mentioned in Theorem 3, we have in Fig. 1 the four translations of a 3-cube with centres $(0,0,0)$, $(0,2,0)$, $(0,1,2)$ and $(2,1,1)$ in a neighborly family, related to the decomposition $K_4 = K_{1,1} + K_{1,2} + K_{1,3}$; Fig. 2 shows it for $K_4 = K_{2,2} + K_{1,2} + K_{1,2}$, with centres $(1,0,0)$, $(1,2,0)$, $(0,1,2)$ and $(2,1,2)$.

It follows that a complete characterization of the decompositions of $K_n$ into $d$ complete bipartite subgraphs is equivalent to a complete characterization of the combinatorial types of neighborly families of $n$ translations of the $d$-cube in $E^d$.

J. Kasem [**2**], a student of M. Perles, has recently been successful in finding small collections of neighborly translations of the $d$-cube in $E^d$, which are maximal in the sense that they cannot be enlarged; he did it by treating some specific types of decompositions of $K_n$ into complete bipartite subgraphs.

We close with the following:

COROLLARY. *The maximum number of neighborly translations of the d-cube in $E^d$ is $d + 1$, for all $d$.*

*Proof*. By Theorems 1 and 3, if there is a neighborly family of $n$ translations of the $d$-cube in $E^d$, then $d \geqslant n - 1$; hence $n \leqslant d + 1$; thus at most $d + 1$ translations of the $d$-cube in $E^d$ can be neighborly.

The construction of a neighborly family of $d + 1$ translations of the $d$-cube in $E^d$ can be easily done inductively, or via the decomposition of $K_{d+1}$ into any $d$ complete bipartite graphs, for example $K_{d+1} = \sum_{i=1}^{d} K_{1,i}$.

**Acknowledgement.** The author wishes to thank N. Linial for his helpful remarks during an early presentation of this subject in Jerusalem, 1982.

### References

**1.** R. L. Graham and H. O. Pollak, On the addressing problem for loop switching, Bell System Tech. J., 50 (1971), 2495–2519; see also "On embedding graphs in squashed cubes", in Graph Theory and its Applications, Y. Alavi et al. ed., Lecture Notes in Math., Springer-Verlag, vol. 303 (1972) 99–110.

**2.** J. Kasem, Ph.D. Thesis, Hebrew University, Jerusalem, 1984.

**3.** G. W. Peck, New proof of a theorem of Graham and Pollak, Discrete Math., 49 (1984) 327–328.

**4.** B. Reznick, P. Tiwari, and D. B. West, Decomposition of product graphs into complete bipartite subgraphs, U. of Illinois, Urbana, 1984.

**5.** H. Tverberg, On the decomposition of $K_n$ into complete bipartite graphs, J. Graph Theory, 6 (1982) 493–494.

**6.** J. Zaks, Bounds of neighborly families of convex polytopes, Geom. Dedicata, 8 (1979) 279–296.

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## SOLUTION OF POLYNOMIALS BY REAL RADICALS*

I. M. ISAACS

*Department of Mathematics, University of Wisconsin, Madison, WI 53706*

Let $F$ be a field and let $f \in F[x]$ be a polynomial. Recall that $f$ is said to be *solvable by radicals* if $f$ splits over some extension field $E \supseteq F$ for which there exists a chain of fields

$$F = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_k = E$$

such that $F_i = F_{i-1}[\alpha_i]$ for $1 \leqslant i \leqslant k$, where $\alpha_i$ has some power lying in $F_{i-1}$. Such an extension field $E$ is called a *repeated radical extension* of $F$.

By the work of Galois, we know that (in characteristic zero) every polynomial of degree $\leqslant 4$ is solvable by radicals, and that over suitable fields (the rational numbers $\mathbb{Q}$, for instance), there exist, for every integer $n \geqslant 5$, some irreducible polynomials which are solvable by radicals and some which are not.

Let us now limit the discussion to the case $F = \mathbb{Q}$. What is the situation with respect to solvability by radicals if we require that we work entirely inside the real numbers $\mathbb{R}$? We shall say that $\alpha \in \mathbb{R}$ is a *real radical element* if it lies in some repeated radical extension $E$ of $\mathbb{Q}$ with $E \subseteq \mathbb{R}$. In other words, $\alpha$ is constructible from the rationals by some combination of multiplications, additions and extractions of real $n$th roots.

A polynomial $f \in \mathbb{Q}[x]$ is *solvable* by *real radicals* if all of its (complex) roots are real radical elements. As we shall see, a polynomial which is solvable by radicals is rarely solvable by real radicals even given that all of its roots are real. More surprisingly, we have the following:

THEOREM. *Suppose $f \in \mathbb{Q}[x]$ is an irreducible polynomial which splits over $\mathbb{R}$. (In other words, its roots are all real.) If $f$ has any root which is a real radical element, then the degree of $f$ is a power of 2 and the Galois group of $f$ over $\mathbb{Q}$ is a 2-group.*

The fact that irreducible cubics are never solvable by real radicals is classical. This phenomenon is sometimes referred to as the "casus irreducibilis". (See the discussion of this in [**2**].)

*Research partially supported by an N.S.F. grant.

A striking consequence of the theorem is the following:

COROLLARY. *Let $f$ be as in the theorem. Then all of the roots of $f$ lie in a repeated square-root extension of $\mathbb{Q}$ and thus they are constructible with compass and straightedge.*

*Proof.* The Galois group, because it is a 2-group, necessarily has a chain of subgroups, each of index 2 in the preceding, which begins with the whole group and ends with the trivial subgroup. By the fundamental theorem of Galois theory, the corresponding fields form a tower of quadratic (and hence square-root) extensions, starting with $\mathbb{Q}$ and ending with a splitting field for $f$. ■

The key step in the proof of the theorem is the following result in which we use the notation $|L: F|$ to denote the degree of the field extension $F \subseteq L$.

PROPOSITION. *Let $F \subseteq E$ be subfields of $\mathbb{R}$ and assume that $E = F[\alpha]$ with $\alpha^n \in F$. If $F \subseteq L \subseteq E$ with $L$ Galois over $F$, then $|L: F| \leqslant 2$.*

The proofs of the proposition and the theorem make heavy use of the theorem on "natural irrationalities" (Theorem 29 of [**1**]). We state this here for the convenience of the reader after a brief review of the relevant notation. If $E$ and $L$ are subfields of some common field, their *compositum* $\langle E, L\rangle$ is the (unique) smallest subfield containing both $E$ and $L$. It is most simply defined as the intersection of all subfields which contain the two given fields. For fields $F \subseteq E$, we write $\mathrm{Gal}(E/F)$ to denote the associated Galois group.

THEOREM (Natural irrationalities). *Let $F$, $E$, and $L$ be subfields of some field $\Omega$ and suppose $E \supseteq F$ is a Galois extension. Let $K = \langle E, L\rangle$, the compositum. Then $K$ is Galois over $L$ and* $\mathrm{Gal}(K/L) \cong \mathrm{Gal}(E/E \cap L)$. *In particular*, $|K: L| = |E: E \cap L|$. (See Fig. 1.) ■

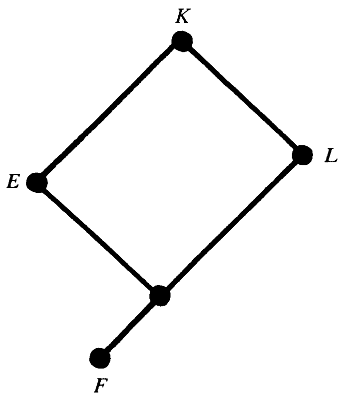



FIG. 1

COROLLARY 1. *In the above situation, assume $|L: F| < \infty$. Then $|K: E| = |L: E \cap L|$.*

*Proof.* Divide both sides of the equation $|K: L| = |E: E \cap L|$ into the number $|K: E \cap L|$. ■

A somewhat more interesting corollary permits us to draw another line in the natural irrationalities lattice diagram.

COROLLARY 2. *Let $F$, $L$, $E$ and $K$ be as before with $E$ Galois over $F$ and $K = \langle E, L\rangle$. Suppose $L \subseteq M \subseteq K$. Then $M = \langle L, E \cap M\rangle$.*

The point here is that by drawing the line segment down from $M$, parallel to $KE$, we define a new node, identified with the field $E \cap M$. The resulting diagram, if interpreted as a lattice diagram, would imply $M = \langle L, E \cap M\rangle$. Corollary 2 asserts that this is a valid inference. Of course, the diagram is not a proof. (See Fig. 2.)

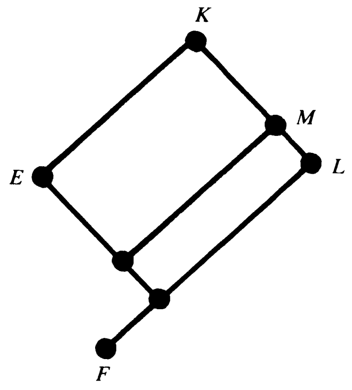



FIG. 2

*Proof of Corollary* 2. Let $M_1 = \langle L, E \cap M \rangle$. It is clear that $M_1 \subseteq M$, $E \cap M_1 = E \cap M$ and $\langle M_1, E \rangle = K = \langle M, E \rangle$. Applying natural irrationalities with $M$ in place of $L$ yields that $|K: M| = |E: E \cap M|$ and similarly, $|K: M_1| = |E: E \cap M_1|$. We conclude that $|K: M| = |K: M_1|$, and since $M_1 \subseteq M$, we have $M_1 = M$ as desired. ■

We will also need the following lemma which is probably well known.

LEMMA. *Let $F \subseteq E$ be fields with $E = F[\alpha]$ where $\alpha^n \in F$, and suppose $F$ contains a primitive nth root of unity. Let $|E: F| = m$. Then*

$$\alpha^m \in F.$$

*Proof*. Let $f \in F[x]$ be the minimal polynomial of $\alpha$ over $F$ so that $\deg(f) = m$. Write $\alpha^n = a \in F$. Then $f$ divides $x^n - a$ in $F[x]$ and so all the roots of $f$ in a splitting field are $n$th roots of $a$. Each root, therefore, has the form $\alpha\delta$ where $\delta$ is some $n$th root of unity. It follows that the constant term of $f$ is equal to $\pm\alpha^m\varepsilon$, where $\varepsilon$ is the product of all the $\delta$'s and so is an $n$th root of unity. Thus $\varepsilon \in F$ and we conclude that $\alpha^m \in F$ as required. ■

*Proof of the proposition*. Adjoin a complex primitive $n$th root of unity to each of $F$ and $E$ to obtain the fields $F^*$ and $E^*$, respectively. Note that $E^* = \langle F^*, E \rangle$ and let $K = \langle F^*, L \rangle$ and $M = E \cap K$ so that we have the lattice diagram as shown. (See Fig. 3.)

Since $F^*$ is a splitting field for the polynomial $x^n - 1$ over $F$, we see that $F^* \supseteq F$ is Galois and in fact $\mathrm{Gal}(F^*/F)$ is abelian. By Corollary 2, $M = \langle L, F^* \cap M \rangle$ and we observe that $L$ is Galois over $F$ by hypothesis and $F^* \cap M$ is Galois over $F$ by the fundamental theorem of Galois theory, since $\mathrm{Gal}(F^*/F)$ is abelian and so $\mathrm{Gal}(F^*/F^* \cap M)$ is a normal subgroup.

Since each of the fields $L$ and $F^* \cap M$ is Galois over $F$, it follows that any $F$-automorphism of any field which contains $\langle L, F^* \cap M \rangle$ leaves each of them (setwise) invariant. Since the compositum (which is equal to $M$) is uniquely determined, it follows that it too is invariant under all these automorphisms and therefore $M$ is normal, and hence Galois over $F$.

By natural irrationalities applied to the situation $\langle F^*, M \rangle = K$ with $F^*$ Galois over $F$, we conclude that $K$ is Galois over $M$. We can therefore apply natural irrationalities to the situation $\langle K, E \rangle = E^*$ and we conclude from Corollary 1 that $|E^* : K| = |E: M|$.

Now write $m = |E^* : K|$. Since $E^* = K[\alpha]$, $\alpha^n \in K$ and $K$ contains a primitive $n$th root of unity, we conclude from the lemma that $\alpha^m \in K$. Since, of course, $\alpha^m \in E$, we see that $\alpha^m \in M$. We claim, in fact, that $M = F[\alpha^m]$.

To see this, write $M_1 = F[\alpha^m]$ and note that $E = M_1[\alpha]$ and $\alpha$ is a root of the polynomial $x^m - \alpha^m \in M_1[x]$. It follows that $|E: M_1| \leqslant m$. However, $M_1 \subseteq M$ and $|E: M| = |E^* : K| = m$, and it follows that $M_1 = M$, as claimed.

Write $\beta = \alpha^m$. The situation now is that $F \subseteq M \subseteq \mathbb{R}$, $M = F[\beta]$, $\beta^n \in F$ and $M$ is Galois

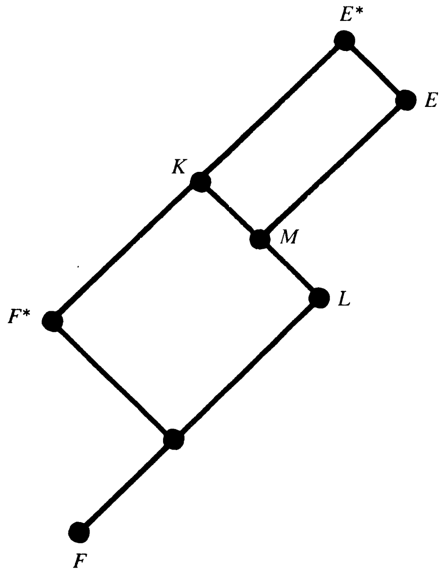



FIG. 3

over $F$. Let $\sigma \in \mathrm{Gal}(M/F)$. Then

$$(\beta^\sigma)^n = (\beta^n)^\sigma = \beta^n$$

and so $\beta^\sigma = \beta\delta$, where $\delta \in M$ is some $n$th root of unity. Since $M \subseteq \mathbb{R}$, however, $\delta = \pm 1$ and thus $\beta^\sigma = \pm\beta$ for every element $\sigma \in \mathrm{Gal}(M/F)$. Since $M = F[\beta]$, $\sigma$ is determined by its action on $\beta$ and it follows that $|\mathrm{Gal}(M/F)| \leqslant 2$.

We now have

$$|L : F| \leqslant |M : F| = |\mathrm{Gal}(M/F)| \leqslant 2. \blacksquare$$

*Proof of the theorem.* Suppose $f \in \mathbb{Q}[x]$ is irreducible and let $S \subseteq \mathbb{R}$ be a splitting field for $f$. Let $G = \mathrm{Gal}(S/\mathbb{Q})$. Our goal is to show that $G$ is a 2-group. Since $\deg f$ divides $|S : \mathbb{Q}| = |G|$, this will yield that $\deg f$ is a power of 2, as required.

Suppose that $G$ is not a 2-group and choose an odd prime divisor $p$ of $|G|$. Let $N \subseteq G$ be the subgroup generated by all subgroups $P \subseteq G$ with $|P| = p$. Since this collection of subgroups is permuted by conjugation by elements of $G$, it follows that $N$ is normal in $G$. Also, by Cauchy's theorem, there do exist subgroups $P \subseteq G$ of order $p$, and hence $N > 1$.

We are given the real radical element $\alpha$ with $f(\alpha) = 0$, and so $\alpha \in S$ and we write $T = \mathbb{Q}[\alpha]$. Let $H = \mathrm{Gal}(S/T)$, so that $H \subseteq G$ is the subgroup corresponding to $T$ via the fundamental theorem of Galois theory. We claim that $N$ cannot be contained in $H$. To see this, let $L = \mathrm{Fix}(N)$, which is the field corresponding to $N$. Then $L$ is normal over $\mathbb{Q}$ since $N \triangleleft G$. Also, $L < S$ because $N > 1$. If $N \subseteq H$, then $L \supseteq T$ and so $\alpha \in L$. Because $f$ is irreducible and $L$ is normal over $\mathbb{Q}$, we conclude that $f$ splits over $L$ and this is a contradiction since $L < S$.

We have established that $N \nsubseteq H$ and thus $P \nsubseteq H$ for some subgroup $P \subseteq G$ of order $p$. Let $F = \mathrm{Fix}(P)$. Since $P \nsubseteq H$, we have $F \nsupseteq T$ and so $\alpha \notin F$.

By our assumption on $\alpha$, there exists a tower of fields

$$\mathbb{Q} = Q_0 \subseteq Q_1 \subseteq \cdots \subseteq Q_m \subseteq \mathbb{R}$$

with $\alpha \in Q_m$ and $Q_i = Q_{i-1}[\gamma_i]$ for some element $\gamma_i$ with a power in $Q_{i-1}$. Construct a new tower of fields starting with $F$ by setting $F_0 = F$ and $F_i = F_{i-1}[\gamma_i]$. Then $Q_i \subseteq F_i \subseteq \mathbb{R}$ for all $i$ and, in particular, $\alpha \in F_m$. Since $\alpha \notin F_0$, we can choose an integer $r > 0$ so that $\alpha \notin F_{r-1}$ and $\alpha \in F_r$.

Now $|S: F| = |P| = p$, a prime, and thus $F$ is a maximal subfield of $S$. Also,

$$F \subseteq F_{r-1} \cap S \subseteq F_r \cap S \subseteq S.$$

Since $\alpha \notin F_{r-1} \cap S$ and $\alpha \in F_r \cap S$, we conclude that $F = F_{r-1} \cap S$ and $F_r \cap S = S$ so that $S \subseteq F_r$.

Write $E = \langle S, F_{r-1} \rangle$. Then $F_{r-1} \subseteq E \subseteq F_r$. Also, $S$ is Galois over $\mathbb{Q}$ and so natural irrationalities applies and we conclude that $|E: F_{r-1}| = |S: F| = p$ and $E$ is Galois over $F_{r-1}$. However, $F_{r-1}[\gamma_r] = F_r \subseteq \mathbb{R}$ and some power of $\gamma_r$ lies in $F_{r-1}$. Our proposition now applies and yields $p = |E: F_{r-1}| \leqslant 2$. This contradiction proves the theorem. (See Fig. 4.) ■

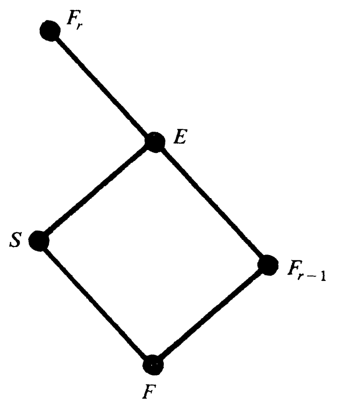



FIG. 4

We close with the observation that solvable polynomials with real roots but which are not solvable by real radicals seem to abound. For example, for any prime $p$, the polynomial

$$f(x) = x^3 - 2px + p$$

has this property. (Note that it is irreducible by the Eisenstein criterion and it has three real roots since $f(0) > 0$ and $f(1) < 0$.) It is amusing to solve this polynomial by Cardan's method to see where nonreal numbers come in.

**References**

1. E. Artin, Galois Theory, Notre Dame Math. Lectures, 1944.
2. B. L. van der Waerden, Modern Algebra (vol. 1, 2nd edition), Ungar, New York, 1953.

## A THEOREM IN COMBINATORICS AND WILSON'S THEOREM

HARAGAURI N. GUPTA
*Department of Mathematics and Statistics, University of Regina, Regina, Canada S4S 0A2*

The purpose of this note is to state and prove a theorem in combinatorics and deduce from it the famous Wilson's theorem as a special case.

THEOREM 1. *Let $B(n,r)$, $1 \leqslant r \leqslant n$, denote the number of distributions of $n$ distinct objects among $r$ ordered boxes, none remaining empty. Then,*

(i) *if $n$ is prime, then $n$ divides $B(n,r)$ for all $r \geqslant 2$,*
(ii) *if $n$ is prime, then $n$ divides $B(n-1,r) + (-1)^r$, for all $r$, $1 \leqslant r \leqslant n-1$.*

*Proof.* We shall not require the explicit formula for $B(n,r)$, but will make do with the following identities I and II, both of which can be established solely by combinatorial arguments:

## Telegraphic Reviews

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, S*(14-16), L*.** A Diary on Information Theory. Alfréd Rényi. Akademiai Kiado, 1984, 192 pp, $22.50. [ISBN: 963-05-3876-8] A sequel to Rényi's well-known Dialogues on Mathematics and Letters on Probability, the "diary" is Rényi's rendering of an imaginary student's reactions to his first university lecture series on information theory, including thoughts on mathematics, pedagogy, science, and styles of learning. Also includes four popular lectures by Rényi on related topics. LAS

**Mathematics Appreciation, S(13-14), L.** Magic Squares of Order Four. Kathleen Ollerenshaw, Hermann Bondi. Royal Society (US Distr: Scholium Intern, 265 Great Neck Road, Great Neck, NY 11021), 1982, 88 pp, $26 (P). Reprinted from Philosophical Transactions of the Royal Society, Volume 326, A1495, October 1982, pages 443-532. By a wholly new method using "Frenicle quads" and "part sums" the authors confirm that there are essentially 880 different normal 4x4 magic squares, a result established about 300 years ago by Frenicle using the method of exhaustion. A 4x4 square is normal if constructed from the integers 1 - 16 (or, equivalently, 0 - 15). Two squares are essentially different if neither is obtainable from the other by rotations and/or reflections. All 880 solutions are given in an appendix. The paper opens with a history of the 4x4 square. JK

**Mathematics Appreciation, S(13-16).** Eine mathematische Reise. Friedrich Wille. Vandenhoeck & Ruprecht, 1984, 119 pp, DM 12.80 (P). [ISBN: 3-525-33503-2] An amusing account of an imaginary trip in which what happens raises mathematical questions from various fields, including geometry, algebra and set theory. JD-B

**Precalculus, S(13-14), P, L.** Algebra, Geometry and Trigonometry in Science, Engineering and Mathematics. M.V. Sweet. Ser. in Math. & Its Applic. Halsted Pr, 1984, 617 pp, $45. [ISBN: 0-470-20102-9] Covers non-calculus topics for first- and second-year university students in Scotland: precalculus topics plus probability, series, vectors, linear algebra and three-dimensional coordinate geometry. Numerous examples and exercises. Would serve as a useful compendium. JNC

**Precalculus, T(13: 1), S.** Intermediate Algebra. L. Murphy Johnson, Arnold R. Steffensen. Scott Foresman, 1985, 552 pp, $24.95. [ISBN: 0-673-15632-X] Text is designed for review of basic concepts and formulas. Includes review questions from previous material covered. Contains several solved examples for each topic. Easy to read. TR

**History, P, L.** An Idiot's Fugitive Essays on Science: Methods, Criticism, Training, Circumstances. C. Truesdell. Springer-Verlag, 1984, xvii + 654 pp, $58. [ISBN: 0-387-90703-3] A miscellany of reprints of essays and reviews, a few never before published, written by Truesdell during the last 30 years. The major parts are reviews and biographies emphasizing the history of mechanics; other essays on aspects of the philosophy of science reflect the "fugitive" Truesdell's curmudgeonly scholarship. A fascinating source of opinion, quotations, and historical information. LAS

**History, L*.** Oeuvres Complètes. Elie Cartan. Springer-Verlag, 1984, $87.10 set [ISBN: 0-387-13629-0] Partie I, xxx + 1356 pp; Partie II, xiii + 1384 pp; Partie III/1, xvi + 992 pp; Partie III/2, xvi + 938 pp. Photocopy of the 1952-1955 Gauthiers-Villars edition. Part I comprises Lie groups, Part II algebra and differential equations, and Part III (in two volumes) differential geometry. Each part contains a complete list of Cartan's work and a frontspiece portrait; the final volume concludes with two obituary Notices by Chern and Chevalley, and by J.H.C. Whitehead. LAS

**Combinatorics, S(18), P.** Progress in Graph Theory. Ed: J. Adrian Bondy, U.S.R. Murty. Academic Pr, 1984, xiii + 539 pp, $59.50. [ISBN: 0-12-114320-1] Proceedings from the Silver Jubilee Conference on Combinatorics held at the University of Waterloo in 1982. Includes major addresses by Bollobas and Thomasson along with forty-one research papers and a section on unsolved problems. JS

**Combinatorics, P.** Enumeration and Design. Ed: David M. Jackson, Scott A. Vanstone. Academic Pr, 1984, x + 317 pp, $39.50. [ISBN: 0-12-379120-0] A collection of papers presented at the conference

celebrating the 25th anniversary (1982) of the University of Waterloo. Among the contributions in enumeration: combinatorial commutative algebra, asymptotic methods, partition theory and graphical enumeration. Among the contributions in combinatorial design: balanced incompleted block designs, partial geometries, difference sets, covering problems and coding. SS

**Number Theory, T(13: 1), S*, L*.** Microchip Mathematics: Number Theory for Computer Users. Keith Devlin. Shiva, 1984, vii + 205 pp, £12.95 (P). [ISBN: 1-85014-047-2] An elementary introduction to selected parts of number theory (primes, congruences, primality testing) with extensive exercises for computer exploration. The author is a regular columnist on computer mathematics for The Guardian. This short typewritten monograph could be used either for self study or for a freshman seminar. LAS

**Number Theory, T(17: 1), S, P*, L.** Lecture Notes in Mathematics-1087: Uniform Distribution of Sequences of Integers in Residue Classes. Władysław Narkiewicz. Springer-Verlag, 1984, viii + 125 pp, $7.50 (P). [ISBN: 0-387-13872-2] These notes present a survey of what is known about the uniform distribution of sequences of integers into residue classes. Includes an excellent collection of exercises and a list of references. CEC

**Linear Algebra, T??(14-15: 1, 2), S, L.** Real Linear Algebra. Antal E. Fekete. Pure & Appl. Math., V. 91. Dekker, 1985, xxi + 426 pp, $47.25. [ISBN: 0-8247-7238-5] Interesting, different, highly geometric approach based on ideas of Steenrod. That the cross product makes $R^3$ a Lie algebra of antisymmetric operators (corresponding to the rotation group) is a major theme. Operator approach, matrices playing secondary role for illustration and calculation. Steenrod's exponential formula for rotations, exponential functor, geometric motivations (especially for $R^3$) predominate. RM

**Group Theory, T(18: 1), S, P.** Representations of General Linear Groups. G.D. James. London Math. Soc. Lect. Note Ser., V. 94. Cambridge U Pr, 1984, xii + 147 pp, $19.95 (P). [ISBN: 0-521-26981-4] An investigation of the unipotent representations of the finite general linear groups and the relationships existing between these and representations for the symmetric groups. Consideration is given both to the "natural characteristic" for the ground field and also to the general case. Bibliography. JS

**Topological Groups, S(18), P.** Pseudogroupes de Lie transitifs, I: Structures principales. C. Albert, P. Molino. Hermann, 1984, 142 pp, 130F (P). [ISBN: 2-7056-5978-1]

**Algebra, P.** Représentations des groupes réductifs sur un corps local. J.-N. Bernstein, et al. Hermann, 1984, 157 pp, 160F (P). [ISBN: 2-7056-5989-7]

**Algebra, S(15-16).** Algebra through Practice: A Collection of Problems in Algebra with Solutions, Books 1, 2 & 3. T.S. Blyth, E.F. Robertson. Cambridge U Pr, 1984, x + 291 pp, $6.95. [ISBN: 0-521-25300-4] A collection of exercises, mostly routine and computational, together with solutions. Organized into three parts: one on sets, relations and mappings; a second on matrices and vector spaces; and the third on groups, rings, and fields. Sample tests for each part. JS

**Algebra, S(16-18), P, L.** Elements of Abstract Algebra. Allan Clark. Dover, 1984, xvii + 206 pp, $6 (P). [ISBN: 0-486-64725-0] A re-issue in paperback of the original book published by Wadsworth in 1971 (TR, October 1972). JS

**Calculus, S*(13).** The Calculus Companion to Accompany Calculus, Second Edition, by Howard Anton. William H. Barker, James E. Ward. Wiley, 1984. Volume 1, $15.95 (P) [ISBN: 0-471-09230-4]; Volume 2, $12.95 (P) [ISBN: 0-471-88614-9]. A hefty collection of self-contained units which correspond to Anton's chapters. A chatty treatment of algebra and trigonometry review, motivation, word problem strategy and detailed computations which marvelously complements the text. No additional problems or proofs. Focuses on deepening rather than expanding the discussion. MA

**Calculus, T*(13-14: 1-3).** Calculus. Harley Flanders. WH Freeman, 1985, xii + 1045 pp, $38.95. [ISBN: 0-7167-1643-7] For the average student. Written in the current mode with emphasis on the intuitive approach. Informal and meant to be non-intimidating. Should be readable by students. Extensive warnings, hints and problem-solving suggestions. For clarity, careful attention has been paid to the diagrams and to the display of mathematical expressions. Graded exercise sets with difficulty ranging from routine to challenging. Derivatives of trigonometric functions are introduced in Chapter 2, of exponential and logarithmic functions in Chapter 4. Study guide, instructor's manual and diskette for Apple computer of useful calculus-related programs written in Pascal are available. JK

**Calculus, T*(13: 1).** Calculus with Applications to Business, Economics, and Social Sciences. Richard Bouldin. Saunders College, 1985, xiv + 524 pp. [ISBN: 0-03-069764-6] Elementary exposition and attractive layout incorporating business, economics and personal finance applications. Each chapter also contains calculator examples and exercises, a career profile, a separate section containing social science applications and review exercises. JNC

**Calculus, T*(13: 2).** Mathematics with Applications to Business, Economics, and Social Sciences. Richard Bouldin. Saunders College, 1985, xvii + 765 pp. [ISBN: 0-03-062164-X] Written in the same style as Bouldin's Calculus (see previous review), this text contains all the material from Calculus except trigonometric functions. Also includes chapters on systems of equations and matrices, linear programming, mathematics of money, sets and counting, probability. JNC

**Real Analysis, P.** Lecture Notes in Mathematics-1089: Measure Theory, Oberwolfach 1983. Ed: D. Kölzow, D. Maharam-Stone. Springer-Verlag, 1984, xiii + 327 pp, $10 (P). [ISBN: 0-387-13874-9] Proceedings of a conference held at Oberwolfach, West Germany. Thirty papers and a problem section are included. PZ

**Complex Analysis, S(18), P.** Lectures on the Theory of Functions of Several Complex Variables. B. Malgrange. Springer-Verlag, 1984, 128 pp, $7.10 (P). [ISBN: 0-387-12875-1] A re-issue of the author's well-known Tata Institute lecture notes from 1958. A few exercises are included. The bibliography is limited and, naturally, dated. PZ

**Differential Equations, S*(16-18), P.** Seminar on Nonlinear Partial Differential Equations. Ed: S.S. Chern. Math. Sci. Res. Inst. Pub., V. 2. Springer-Verlag, 1984, 373 pp, $24. [ISBN: 0-387-96079-1] Eighteen well written research-expository essays whose topics vary widely in scope and level of technicality, for instance: P.D. Lax on shock waves; R. Melrose on propagation of singularities; J.C. Polking on removable singularities; and A. Weinstein on plasma physics. Every reader with some affinity for the subject should find in this collection a few fascinating papers. YN

**Differential Equations, P.** Singularities and Groups in Bifurcation Theory, Volume I. Martin Golubitsky, David G. Schaeffer. Appl. Math. Sci., No. 51. Springer-Verlag, 1985, xvii + 463 pp, $38. [ISBN: 0-387-90999-0] Applies techniques from singularity theory, especially unfolding theory and classification theory, to bifurcation problems. Suitable for applied scientists as well as for mathematicians. Attractively illustrated; many examples and exercises; case studies. Second volume will concern bifurcation problems in several state variables. DA

**Differential Equations, T*(14-15: 1, 2).** Elementary Differential Equations with Applications. C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1985, xvii + 664 pp, $32.95. [ISBN: 0-13-254129-7-01] Very attractive, carefully-written but readable, student-friendly treatment of standard fare. Flexibility permits use with students with different backgrounds and in courses with different emphases. Linear algebra needed for systems is complete and self-contained. BASIC programs for numerical methods require little prior programming experience. Numerous concrete realistic worked-out examples. Exercises and problems range in difficulty from routine to challenging. Good illustrations. JK

**Differential Equations, T(16: 1), S, L*.** Linear Turning Point Theory. Wolfgang Wasow. Appl. Math. Sci., No. 54. Springer-Verlag, 1985, ix + 246 pp, $38. [ISBN: 0-387-96046-5] "Linear turning points" are singularities, in the complex plane, of linear analytic ordinary differential equations. Begins with an historical account of the subject. Then demonstrates the classical and modern methods for studying the behavior of solutions around singularities: asymptotic power series and expansions, parameter shearing, doubly asymptotic expansions, etc. Ends with an appendix on holomorphic similarities of holomorphic matrices to Jordan's and Arnold's forms. Without exercises, but well written. YN

**Partial Differential Equations, P.** Navier-Stokes Equations: Theory and Numerical Analysis, Third (Revised) Edition. Roger Temam. Stud. in Math. & Its Applic., V. 2. Elsevier Sci, 1984, xii + 526 pp, $35 (P); $84.50. [ISBN: 0-444-85308-1; 0-444-85307-3] First treats the linearized stationary case, then the nonlinear stationary case, finally the full nonlinear time-dependent case. Contains an updating of the bibliography of earlier editions and a description of some directions in which there has been recent progress. (First Edition, TR, December 1978; Revised Edition, TR, November 1981) DA

**Functional Analysis, T(15-17: 2), L.** Applied Functional Analysis. D.H. Griffel. Ser. in Math. & Its Applic. Halsted Pr, 1981, 386 pp, $27.95 (P). [ISBN: 0-470-27196-5] A comprehensive view of functional analysis, covering and proving most of the well-known results in distribution theory and Green's functions, Banach spaces and fixed point theorems, and operators in Hilbert space. Definitions are well illustrated with examples, theorems are carefully motivated, and proofs are detailed and readable. MU

**Analysis, T(18: 2), P, L.** Topology and Analysis: The Atiyah-Singer Index Formula and Gauge-Theoretic Physics. B. Booss, D.D. Bleecker. Transl: D.D. Bleecker, A. Mader. Universitext. Springer-Verlag, 1985, xvi + 451 pp, $34 (P). [ISBN: 0-387-96112-7] A well-motivated introduction to the Atiyah-Singer Index theorem replete with historical comments, alternative routes (cobordism, imbedding, heat equation), helpful sketches, extensive references, and wide-ranging connections to many parts of mathematics and physics. Includes necessary introductions to operators, manifolds, K-theory, and differential equations. A well done thorough survey of the "deepest theorem" in modern analysis. LAS

**Analysis, T(17-18: 1, 2), S, P.** A Treatise on Generating Functions. H.M. Srivastava, H.L. Manocha. Ser. in Math. & Its Applic. Halsted Pr, 1984, 569 pp, $89.95. [ISBN: 0-470-20010-3] A systemic introduction to "methods of obtaining linear, bilinear, bilateral or mixed multilateral generating functions for a fairly wide variety of sequences of special functions." FLW

**Algebraic Geometry, T(18: 2), P*.** Geometry of Algebraic Curves, Volume I. E. Arbarello, et al. Grund. der math. Wissenschaften, B. 267. Springer-Verlag, 1985, xvi + 386 pp, $34. [ISBN: 0-387-90997-4] Contains much advanced material on algebraic curves on compact Riemann surfaces. The material goes much beyond the standard results like the Riemann-Roch theorem. Some topics are varieties of special divisors and the geometry of the theta divisor. LN

**Differential Geometry, P.** Differential Geometry and Complex Analysis. Ed: I. Chavel, H.M. Farkas. Springer-Verlag, 1985, xiii + 222 pp, $32. [ISBN: 0-387-13543-X] A collection of papers in differential geometry and complex analysis, in memory of H.E. Rauch. Three historical papers survey and assess Rauch's contributions to differential geometry, function theory, and theta functions. PZ

**Optimization, S(17-18), P.** Minimization Methods for Non-Differentiable Functions. N.Z. Shor. Transl: K.C. Kiwiel, A. Ruszczyński. Ser. in Comput. Math., V. 3. Springer-Verlag, 1985, viii + 162 pp, $30. [ISBN: 0-387-12763-1] Theoretical treatment of methods for minimization of non-smooth functions, using generalized gradient sets, accelerated subgradient methods based on space dilations. Applications to important mathematical programming problems. RM

**Optimization, T(14-16), S.** Linear Programming with BASIC and FORTRAN. Carvel S. Wolfe. Reston, 1985, x + 293 pp, (P). [ISBN: 0-8359-4082-9] Readable yet complete introduction to linear and integer programming. Simplex method, duality, parametric and cutting plane techniques, matrix games, transportation problems. Nice examples, many exercises. Discusses subroutines for pivoting. Updated version of 1973 FORTRAN edition. (TR, August-September 1973) RM

**Control Theory, T*(16: 1), S*, P*, L*.** Linear Multivariable Control: A Geometric Approach, Third Edition. W. Murray Wonham. Applic. of Math., No. 10. Springer-Verlag, 1985, xvi + 334 pp, $42. [ISBN: 0-387-96071-6] Mathematical treatment of controllability, observability, tracking, disturbance decoupling, and quadratic optimization, preceded by a review of prerequisites in abstract linear algebra. Uses linear algebra abundantly, and successfully, to avoid cumbersome matrix computations. Written carefully enough to be understandable by senior undergraduates in mathematics or engineering; includes exercises. Appropriate choice for an applied course that nicely pulls together linear algebra, ordinary differential equations, holomorphic functions, and Laplace transforms. YN

**Probability, P.** Lecture Notes in Mathematics-1095: Stochastic Analysis and Applications. Ed: A. Truman, D. Williams. Springer-Verlag, 1984, v + 199 pp, $11 (P). [ISBN: 0-387-13891-9] Proceedings of the International Conference held in Swansea, Wales, in April 1983. Contains 13 papers, most from the conference, the rest more recent research results by the Swansea school. Applications are mainly in areas related to mathematical physics. RSK

**Probability, T(18: 1, 2), S, P.** Markov Chains, Revised Edition. D. Revuz. Math. Lib., V. 11. Elsevier Sci, 1984, xi + 374 pp, $57.75. [ISBN: 0-444-86400-8] This new edition contains material on zero-one laws and sub-additive ergodic theory. Assumes knowledge of measure theoretic probability, point set topology, topological groups, and Fourier transforms (First Edition, TR, February 1976). FLW

**Statistics, T(18: 1), P.** Multivariate Calculation: Use of the Continuous Groups. Roger H. Farrell. Ser. in Stat. Springer-Verlag, 1985, xvi + 376 pp, $42. [ISBN: 0-387-96049-X] Expansion of the author's 1976 paperback, #520 in the Springer-Verlag Lecture Notes in Mathematics series, entitled Techniques of Multivariate Calculation (TR, May 1977). Integrated treatment of various mathematical techniques for finding probability density functions, together with required background material on topics such as null sets, exterior algebra, symmetric functions, and consequences of the uniqueness theorem for Haar measures. RSK

**Statistics, P.** Multivariate Analysis: Methods and Applications. William R. Dillon, Matthew Goldstein. Wiley, 1984, xii + 587 pp, $37.95. [ISBN: 0-471-08317-8] In the Wiley Series in Probability and Mathematical Statistics. Non-theoretical but thorough presentation, illustrated with real data, designed for users who want more than a "cookbook." Includes chapters on cross-classified categorical data, latent structure analysis, and linear structural relations, in addition to the more standard topics. RSK

**Statistics, T(13-14: 1, 2), S.** Statistics & Probability & Their Applications. Patrick Brockett, Arnold Levine. Saunders College, 1984, xvi + 576 pp. [ISBN: 0-03-053406-2] Presupposes only high school mathematics. The usual topics plus more emphasis on data collection, and survey questionnaire design. FLW

**Statistics, P,** Lecture Notes in Economics and Mathematical Systems-237: Misspecification Analysis. Ed: Theo K. Dijkstra. Springer-Verlag, 1984, v + 129 pp, $9.50 (P). [ISBN: 0-387-13893-5] Proceedings of a Workshop held in Groningen, The Netherlands, December 15-16, 1983. Contains eight papers covering a wide range of topics. RSK

**Statistics, S.** An Introduction to Linear Regression and Correlation, Second Edition. Allen L. Edwards. WH Freeman, 1984, xv + 206 pp, $19.95 (P). [ISBN: 0-7167-1594-5] Claims to require only a "working knowledge of elementary algebra." Discusses equations for lines, bivariate regression, several sorts of correlation coefficients and tests for them, multiple regression, matrix algebra, dummy variables, and orthogonal polynomials. The examples are from psychology. FLW

**Statistics, S(15-18), P*, L.** W.G. Cochran's Impact on Statistics. Ed: Poduri S.R.S. Rao, Joseph Sedransk. Wiley, 1984, xvii + 431 pp, $41.95. [ISBN: 0-471-09912-0] In the Wiley Series in Probability and Mathematical Statistics. Interesting collection of 20 articles, written by former students, colleagues and friends, examining Cochran's great influence on both theoretical and applied statistics. Topics discussed are sampling and observational studies, public health and biometrics, categorical data, analysis of variance and design of experiments, variance components, some philosophical issues, and theoretical developments. Valuable reading for those interested in the history

of mathematical statistics. RSK

**Statistics, P*.** Jack Carl Kiefer, Collected Papers. Ed: Lawrence D. Brown, et al. Springer-Verlag, 1985. I: Statistical Inference and Probability (1951-1963), xxxi + 502 pp, $42 [ISBN: 0-387-96003-1]; II: Statistical Inference and Probability (1964-1984), xiii + 590 pp, $42 [ISBN: 0-387-96003-1]; III: Design of Experiments, xxiv + 718 pp, $39. [ISBN: 0-387-96004-X] Includes all of the more than 100 papers Kiefer (1924-1981) wrote, together with brief commentaries on many of them (particularly in the area of design of experiments). Also contains some short articles on his life and work, and a complete bibliography. RSK

**Computer Literacy, S, P, L.** Starting with UNIX. P.J. Brown. Addison-Wesley, 1984, xiv + 221 pp, $12.95 (P). [ISBN: 0-201-10924-7] An excellent, well-written introduction for beginners, moving from UNIX principles (files, directories, pipes, shell) to specific commands and applications: shell, editing ("ed", "grep"), mail, errors, document preparation, programming (including an introduction to C). LAS

**Computer Literacy, S(13).** The Secret Guide to Computers. Russ Walter. Birkhauser Boston, 1984, 349 pp, $14.95 (P). [ISBN: 0-8176-3190-9] A chaotic compendium of detail for computer-phobic users of personal computers, emphasizing common problems, simple BASIC, and popular jargon. Very well written, but the unimaginative type script pages are very difficult to scan and read. LAS

**Computer Programming, S(13-18), P.** Mastering Apple BASIC: II Plus, IIe, IIc. John J. DiElsi, Elaine S. Grossman, John P. Tucciarone. Apple Programming Ser. Holt, Rinehart & Winston, 1984, xiii + 454 pp. [ISBN: 0-03-071173-8] Text is concise and to the point. Contains many programs to illustrate various techniques. Includes several sorting routines and iteration structures. A good all-purpose reference. TR

**Computer Programming, S(16-17), P*, L.** Foundations of Logic Programming. J.W. Lloyd. Symbolic Computation. Springer-Verlag, 1984, x + 124 pp, $17. [ISBN: 0-387-13299-6] The mathematical theory for using logic as a programming language. Introduces logic programming. Establishes results on fixed points, resolution, completeness, independence, negation and concurrency. Presumes logic, PROLOG, induction and metric spaces. RWN

**Computer Science, S(16-17), P.** The Evolution of Programs. Nachum Dershowitz. Progress in Comp. Sci., No. 5. Birkhauser Boston, 1983, 357 pp, $16.95 (P); $24.95. [ISBN: 0-8176-3156-9] Formal methods for manipulating programs for debugging, modification to meet new specifications, abstraction to produce program schemata. Synthesis rules for generating code from specifications, annotation rules to generate assertions and facilitate analysis. Goal to include semi-automated tools in programming environments. RM

**Computer Science, S, P.** Fundamentals of Human-Computer Interaction. Ed: Andrew Monk. Comp. & People Ser. Academic Pr, 1984, xvii + 293 pp, $26.50. [ISBN: 0-12-504580-8] Recently there has been a great deal of research in studying the interface between a human and a computer system. The point of this research has been to learn how to maximize the productivity of the user while minimizing the likelihood of making errors. The areas of research have included screen characteristics, use of speech and graphics interfaces, design of command languages, and physical workstation layout. This text summarizes the current research work and results which have been obtained in this field. MS

**Computer Science, P.** The Measurement of Visual Motion. Ellen Catherine Hildreth. MIT Pr, 1984, 241 pp, $30. [ISBN: 0-262-08143-1] ACM Distinguished Doctoral Dissertation. Deals with measurement of visual motion and perceptual studies of motion. FA

**Computer Science, P.** Lecture Notes in Computer Science-181: Foundations of Software Technology and Theoretical Computer Science. Ed: Mathai Joseph, Rudrapatna Shyamasundar. Springer-Verlag, 1984, viii + 468 pp, $19.50 (P). [ISBN: 0-387-13883-8] Proceedings of a December 1984 conference, the fourth in a series on the same topic, held at Bangalore, India. LAS

**Computer Science, S(18).** Algorithm-Structured Computer Arrays and Networks: Architectures and Processes for Images, Percepts, Models, Information. Leonard Uhr. Comp. Sci. & Appl. Math. Academic Pr, 1984, xxiii + 414 pp, $34.50. [ISBN: 0-12-706960-7] A wide-ranging survey of algorithmic and architectural issues involved in the design of multi-computer systems, networks, parallel processors, arrays, and VLSI. Emphasizes applications that are intrinsically two- and three-dimensional such as image processing, meteorological research, and modelling the human brain. LAS

**Computer Science, T(15-17: 1), S, L.** Introductory Theory of Computer Science. E.V. Krishnamurthy. Springer-Verlag, 1984, xii + 211 pp, $15 (P). [ISBN: 0-387-91255-X] Thorough yet informal introduction to theoretical computer science with emphasis on relationships between theory and practice. Algorithms, machines, recursive functions and computability, grammars and production rules, program correctness, complexity. Too few exercises. RM

**Computer Science, S(18).** Eurographics Tutorials '83. Ed: Paul J.W. ten Hagen. Springer-Verlag, 1984, 425 pp, $34.50. [ISBN: 0-387-13644-4] Introduces computer graphics, covering interactive techniques and information on graphical techniques. Describes a graphics standard (GKS, Graphical Kernel System) as well as experiences in its implementation. Three-dimensional models are covered using GKS. FA

**Applications (Artificial Intelligence), P*.** Catalogue of Artificial Intelligence Tools. Ed: Alan Bundy. Symbolic Computation. Springer-Verlag, 1984, xxv + 150 pp, $17 (P). [ISBN: 0-387-13938-9] A listing of 264 artificial intelligence portable software packages. Most entries include a "telegraphic" review, references and information on availability. Well indexed and cross-referenced. A more thorough edition is accessible on-line in the UK. RWN

**Applications (Artificial Intelligence), S(16), P.** The AI Business: The Commercial Uses of Artificial Intelligence. Ed: Patrick H. Winston, Karen A. Prendergast. MIT Pr, 1984, 324 pp, $15.95. [ISBN: 0-262-23117-4] Edited transcripts of lectures and discussions at a 1983 MIT conference on the commercial potential of artificial intelligence. Academic researchers, financial supporters, industrial R&D and problem solvers discuss expert systems, robotics, intelligent programming environments and industrial applications. RM

**Applications (Biology), S(16-18), P.** An Introduction to the Application of Nonnegative Matrices to Biological Systems. Joan M. Geramita, Norman J. Pullman. Papers in Pure & Appl. Math., No. 68. Queen's U, 1984, vi + 96 pp, (P). An informal presentation (without proofs) of some of the linear algebra and matrix techniques used in applications to biological systems. The discussion is confined to nonnegative transition matrices, considering the alternatives of nilpotence or the existence of positive characteristic values. There are separate chapters for steady matrices and cyclic matrices together with a subsequent chapter on stochastic matrices. Annotated bibliography, citation index. JS

**Applications (Engineering), P, L.** Computational Methods for Fluid Flow. Roger Peyret, Thomas D. Taylor. Ser. in Comp. Physics. Springer-Verlag, 1983, x + 358 pp, $22 (P). [ISBN: 0-387-13851-X] Paperback edition of the 1983 original edition (TR, March 1983). A state-of-the-art survey, in three parts: fundamental numerical methods, incompressible flows, and compressible flows. LAS

**Applications (Physics), P.** J.C. Maxwell, The Sesquicentennial Symposium: New Vistas in Mathematics, Science and Technology. M.S. Berger. Elsevier Sci, 1984, xiv + 279 pp, $55. [ISBN: 0-444-86707-4] Papers on nonlinear phenomena, cosmology, geometry, and Maxwell's equations by such experts as Freeman Dyson, Roger Penrose, Chen Ning Yang, and Irving Segal, presented at an October 1981 symposium at the University of Massachusetts at Amherst in honor of the 150th anniversary of the birth of James Clark Maxwell. LAS

**Applications (Physics), T(17-18: 1, 2), P*.** Non-Linear Elastic Deformations. R.W. Ogden. Ser. in Math. & Its Applic. Halsted Pr, 1984, xv + 532 pp, $95. [ISBN: 0-470-27508-1] Research monograph with problems. Suitable for a graduate course. Applications of the theory of non-linear elasticity to the solution of boundary-value problems and the analysis of the mechanical properties of solid materials capable of large elastic deformations in a purely isothermal setting without reference to thermodynamics. Some previously unpublished results and new approaches to old problems. References throughout. JK

**Applications (Physics), P.** University of Strathclyde Seminars in Applied Mathematical Analysis: Classical Scattering. Ed: G.F. Roach. Math. Ser., No. 9. Shiva (US Distr: Birkhauser Boston), 1984, vi + 160 pp, £8.50 (P). [ISBN: 0-906812-99-2] Proceedings of a September 1984 international conference held at the University of Strathclyde, Scotland. 15 papers on mathematical physics of scattering phenomena are included. PZ

**Applications (Physics), S(13-16), L.** The Creation of Matter: The Universe From Beginning to End. Harald Fritzsch. Transl: Jean Steinberg. Basic Books, 1984, x + 307 pp, $19.95. [ISBN: 0-465-01446-1] A popular yet authoritative account of the origin and progress of the universe, set in a serious context of human values. Physicist Fritzsch explains the structure of matter as a way of probing the "how" and the "why" of the Big Bang. A lucid, inviting exposition of contemporary cosmology. LAS

**Applications (Physics), T(18: 1), S, P.** The Mathematical Theory of Turbulence. M.M. Stanišić. Universitext. Springer-Verlag, 1985, xvi + 429 pp, $29 (P). [ISBN: 0-387-96107-0] This text is composed of two parts: the first deals with the classical approach to turbulence in conjunction with the semi-empirical methods of Prandtl, T.I. Tayor and von Kármán. A generalized law concerning eddy viscosity is developed in detail, the equations of the turbulent boundary layer are discussed, and a new approach to their solution is indicated. Part two deals with statistical theories of turbulence. MU

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

## Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

### Allegheny Mountain Section

The spring meeting of the Allegheny Mountain Section was held at the campus of the University of Pittsburgh at Johnstown on April 26-27, 1985.

Invited Addresses:

"The Covering Radius of Codes," by Neil J.A. Sloan, AT&T Bell Laboratories.
"Discrete Mathematics in the Curriculum," by Martha Siegel, Towson State University.
"Principles of Microcomputer Graphics," by Roy E. Myers, Pennsylvania State University at New Kensington.
"The Bieberbach Conjecture: Solution of a Seventy Year Old Problem," by William E. Kirwan, University of Maryland.

Short Presentations:

"A Non-trivial Example of Uncorrelated Dependent Variables," by Douglas H. Frank, Indiana University of Pennsylvania.
"The Maximal Planar Subgraph of a Graph," by Frank X. Hiergeist, West Virginia University.
"Convergence Is Good, Divergence Is Bad. Or Is It?" by Richard McDermot, Allegheny College.
"Raise High the Root Beam Carpenter and Trigonometric Factoring--An Application," by James E. Hall, Westminster College.
"Packing Points in an Equilateral Triangle," by Ron Harrell, Allegheny College.

Student Papers:

"Game Theory Models for Evolution," by Dennis McDermot, Allegheny College.
"Solving Certain Systems of Differential Equations Using Symbolic Programming," by Kevin Treu, Allegheny College.
"Nonstandard Analysis: How and Why," by Chris Oliver, Allegheny College.
"A Discriminating Journey Through the Land of Curves of Constant Width," by Richard Fenrich, Allegheny College.

### Iowa Section

The seventy-second annual meeting of the Iowa Section was held on April 12-13, 1985 at Drake University at Des Moines, Iowa. There were 72 registrants.

Invited Addresses:

"Here We Go Again: Another Challenge to the Simplex Method," by A.M. Fink, Iowa State University.
"The Bieberbach Conjecture: Some History and Related Problems," by Elgin Johnston, Iowa State University.
"Integral Equations of the First Kind and Ill-posed Problems," by Charles Groetsch, University of Cincinnati.

Short Presentations:

"Bell's Inequality and Einstein's Conception of Reality," by Alan Macdonald, Luther College.
"Numerical Solution of Second Kind Integral Equations," by Glenn Lueke, Iowa State University.
"The Numerical Evaluation of Particular Solutions for Poisson's Equation," by Kendall Atkinson, University of Iowa.
"Numerical Solution of the Heat Equation by the Method of Heat Potentials," by Daniel Willis, University of Iowa.
"Weighted Finite Elements for the Stamp Problem," by A. Bogomolny, University of Iowa.
"Affine Invariant Conditions for the Monotone Convergence of Newton's Method and Applications," by Florian Potra, University of Iowa.
"Endospectral Graphs," by Milan Randic and Alex Kleiner, Drake University.
"Factoring the Characteristic Polynomial," by Bernie Baker, Alex Kleiner, and Milan Randic, Drake University.
"Catalan Numbers and the Pascal Triangle," by Milan Randic, Bernie Baker, and Patsy Milton, Drake University.
"Preliminary Model to Improve Estimation of Prevalence of Malaria," by M. Anne Dow, Maharishi International University.
"Universal Objects: A Unifying Structure in Mathematics," by Catherine Gorini Wadsworth, Maharishi International University.
"Progress Report--High School Teacher Certification in Mathematics and Computer Science in Iowa," by Lynn Olson, Wartburg College; Marilyn Zweng, University of Iowa.
"The MAA Team Project," by Thomas Iverson, Central College.

Student Papers:

"On Spectral Properties of Graphs," by Ruth Gornet, Drake University.
"Is There a Time When One Is More Likely To Have a Heart Attack?" by Steven Reeves, Luther College.
"On Factoring the Characteristic Polynomial," by Nick Street, Drake University.
"Rubik's Cube: Tactile Group Theory," by Thomas Rosholt, Luther College.
"On the Characteristic Polynomials of Graphs with Bridges," by Scott Rothfus, Drake University.
"An Algorithm for Generating Permutations with Applications to High School Courses in Computer Simulation or Finite Mathematics," by Janice Mansheim, Emmanuel Maou, and Phyllis Schindler, Iowa Wesleyan College.

Panel Discussion:

"Discrete vs. Continuous Mathematics in the Undergraduate Curriculum," by Donald Alton, University of Iowa; Larry Naylor, Drake University; Zorabi Honargohar, Morningside College; Lynn Olson, Wartburg College.

## Maryland-D.C.-Virginia Section

The spring meeting of the Maryland-D.C.-Virginia Section was held at Hollins College in Roanoke, Virginia on April 25, 1985. 135 persons attended the meeting.

Invited Addresses:

"How to Factor a Number," by Carl Pomerance, University of Georgia.

Short Presentations:

"True BASIC: A New Programming Language for the IBM," by Howard Penn, U.S. Naval Academy.
"A Property of the Unit Digits of the Fibonacci Sequence," by Herta Freitag, Hollins College.
"An Illustration of Computer Learning," by Bob Lewand, Goucher College.
"A Geometric Treatment of the Concept of Means," by William Sanders, James Madison University.
"Partial Fractions: Classical, Efficient Method--Now Complete," by Donald R. Peeples, Mary Washington College.
"The Mathematical Competition of Modeling--One School's Experience," by Alice T. Williams, Southern Seminary Junior College.
"My Ten Years Last Summer at IFRICS (Institute for Retraining in Computer Science)," by Frederic Gooding, Goucher College.
"Results of the 1985 Math Competition in Modeling," by Ben Fusaro, Salisbury State.
"Groups of Order $q^3,p$," by James Alonso, University of the District of Columbia.
"Estimating Tornado Windspeed Probabilities: Empirical, Theoretical, and Subjective Perspectives," by Bob Abbey, Office of Naval Research; Mary Kay Abbey, Montgomery College.
"Geometry's Cinderella," by C.J. Maloney, National Institute of Health.
"On Fermat d-Pseudoprimes," by Lawrence Somer, George Washington University.
"Integer Solution of $c^2 = a^2 + b^2 - 2ab(p/q)$ and Nice Law of Cosine Problems," by Wilbur Hildebrand, Montgomery College.
"Pell-like Equations," by William Wardlaw, U.S. Naval Academy.
"Independent Weightings in the Ball Weighing/Fake Coin Problem," by Ray Hancock, Emory and Henry College.
"A Non-elegant Way of Expressing the nth Power of the Golden Ratio," by Herta T. Freitag, Hollins College.

Student Papers:

"Probabilistic Analysis of a Numerical Method for Finding Zeros of a Function," by Arthur Benjamin, Johns Hopkins University.
"The Experience and the Product of a COMAP Winner--Modelling in Animal Populations," by John Kent, Dan McCaffey, Mike Caulfield, and John August, Mount Saint Mary's College.
"Some Results on the Period and the Restricted Period of the Fibonacci Series Mod-P," by Anna Werner, American University.
"The Independence of Certain Ascoli Theorems," by David A. Lamb, George Mason University.
"Some New Results Regarding Primitive Roots Modulo a Prime," by Steve Bonner, American University.

Workshop:

"Writing Mathematics Programs for the IBM-PC," by David Schneider, University of Maryland.

## Southeastern Section

The sixth-fourth annual spring meeting of the Southeastern Section met at Wake Forest University, Winston-Salem, North Carolina on April 12-13, 1985. 326 persons attended the meeting.

Invited Lectures:

"Solutions of an Inverse Problem for Fractals and Other Sets," by Michael Barnsley, Georgia Institute of Technology.
"Geometry--A Lot's New and a Lot's Renewed," by John Kenelly, Clemson University.
"Computing and the Changing Nature of College Mathematics," by Donald L. Kreider, Dartmouth College.

Short Presentations:

"Conditionally Convergent Integrals of Convolution Products," by Jonathan Lewin, Kennesaw College.
"Sequence of Iterates for Quasi-nonexpansive Mappings in Locally Convex Spaces," by K.L. Singh, Fayetteville State University.
"On Certain Commutation Relations in a Universal Enveloping Algebra," by Kwok Chi Wong, Fayetteville State University.
"Fixed and Common Fixed Points in Convex Metric Spaces," by Larry Harris, Fayetteville State University.
"A Byte of Simpson's Apple," by Jerry Currence, University of South Carolina at Lancaster.
"Software Tools for Education," by Chantal M. Shafroth and James M. Shoaf, North Carolina Central University.
"Continuous Mappings in Computer Graphics," by Lyndell Kerley, East Tennessee State University.
"Investigating Pseudoconics: An Opportunity for Students to Explore and Discover New Mathematics," by F. Alexander Norman, University of Georgia.
"On Solving Radical Equations," by Bill Bompart, Augusta College.
"Partial Fractions: Classical, Efficient Method Now Complete," by Donald R. Peeples, Mary Washington College.
"An Elementary Proof of Descartes' Rule of Signs," by Curtis Herink, Mercer University.
"A Modest Experiment in Mathematics Education," by David A. Smith, Duke University and Benedict College.
"A Sparse, Banded Matrix: What Is Its Null Space?" by Steve Howard, Wake Forest University.
"Weakly Closed Functions and Hausdorff Spaces," by David A. Rose, Francis Marion College.
"What Every Statistician Should (But Probably Doesn't) Know About Levels of Measurement," by R.W. Hoyer, James Madison University.
"Some Popular Explanation of Slow-fast Cycles, Non-standard Models, and Duck Hunting Theorems," by Vadim Komkov, Winthrop College.
"Report on Usage of MAA Multi-media Applied Mathematics (TEAM)," by Gil Proctor, Clemson University.
"Constructing the Good Multiple Choice Item: Easier Said Than Done," by Linda Boyd, DeKalb Community College.
"College Board Achievement Tests in Mathematics," by Rose C. Hamm, The College of Charleston.
"What's Happening in Mathematics Education," by Virginia Joyner, East Carolina University.
"Mathematics Education in the USSR--A Very Recent Up-date (March 1985)," by James G. Ware, University of Tennessee at Chattanooga.
"Approximating Polygons for Space Filling Curves," by Hans Sagan, North Carolina State University.
"On the Number of Honest Permutations," by Albert F. Gilman, III, and Joseph B. Klerlein, Western Carolina University.
"An Application of the Calculus of Finite Differences to Pascal's Triangle," by Benjamin G. Klein, Davidson College.
"Error in Padé Approximation," by Edward J. Allen, University of North Carolina at Asheville.
"Some Improper Integrals are Only Teasing," by John V. Baxley, Wake Forest University.
"Differentiability: A Paradoxically Sufficient Condition," by Irl C. Bivens, Davidson College.
"Basic Orlicz Space Theory," by Arnold S. Goldstein, Savannah, Georgia.
"Divisibility by Primes," by Reginald Mazeres, Tennessee Technological University.
"A Geometric Portrait for the Parametric Representation of the Pythagorean Triples," by Alvin Tirman, East Tennessee State University.
"Squareful Numbers in Arithmetical Progressions," by Richard L. Robinson, Wofford College.
"The Remaining Continued Fractions in Ramanujan's Second Notebook," by Robert L. Lamphere, Francis Marion College.

Student Papers:

"On Cryptography and the Factorization of Large Integers," by Sebrina Bragdon, Spelman College.
"The MOG: A House of Bricks," by Julie Morrisett, Davidson College.

## North Central Section

The spring meeting of the North Central Section was held at the College of St. Benedict, St. Joseph, Minnesota on April 26-27, 1985. There were 104 registrants.

Invited Addresses:

"The Mathematical Sciences: Role and Expectations by the Scientific and Technical Communities," by Ettore Infante, University of Minnesota.
"Renewing Collegiate Mathematics," by Lynn A. Steen, St. Olaf College.

Short Presentations:

* "An Elementary Derivation of $\prod_{n=1}^{\infty} \cos(x/2^n) = \sin(x)/x$," by James Harper, St. Cloud State University.
* "Viscous Flow on a Rotating Disk," by Jeffrey Rosoff, Gustavus Adolphus College.
"ICME-V," by Francis Hatfield, Mankato State University.
* "Generalized Circular Functions," by Douglas Dunham, University of Minnesota at Duluth.
"Teaching of Experiential Applied Mathematics," by Michael Tangredi, College of St. Benedict.
"Simpson's Paradox and the Open Mapping Theorem," by John Kemper, College of St. Thomas.
* "Existence of Self-Reproducing Programs," by Gary Rockswold, Mankato State University.
"Regarding the Bieberbach Conjecture," by Paul Zorn, St. Olaf College.

## Illinois Section

The sixty-fourth annual meeting of the Illinois Section convened at the University of Illinois, Champaign-Urbana on May 3-4, 1985 with approximately sixty members in attendance.

Invited Addresses:

"The Algebraic Structure of Non-standard Models of Arithmetic," by Angus Macintyre, Yale University.
"Mathematics Curriculum of a Comprehensive Two-year College," by Dale Ewen, Parkland College.
"The Teaching and Learning of Mathematics in U.S. High Schools From an International Point of View," by Ken Travers, University of Illinois.
"PME, RCDPM, and ICME-International Mathematics Conferences in Australia, August 1984," by Carole Bauer, Triton College.
"Computer Software for College Algebra and Calculus," by Larry Dornhoff, University of Illinois.
"The Role of Discrete Mathematics in the First Two Years of the Undergraduate Program," by Al Otto, Illinois State University.
"M.C. Escher as Pioneer Mathematician," by Doris Schattschneider, Moravian College.
"The B.A. in Mathematics--What Should It Signify?" by Heini Halberstam, University of Illinois.

Mini-Course:

"Computer Graphics," by Joan Wyzkoski, Bradley University.

Professor Jon Laible of Eastern Illinois University was presented the Distinguished Service Award of the Section.

## Kentucky Section

The sixty-eighth annual meeting of the Kentucky Section was held on April 19-20, 1985 at Eastern Kentucky University at Richmond, Kentucky. There were 109 registrants.

Invited Addresses:

"Some Observations on Mathematics and Mathematicians," by Ivan Niven, University of Oregon.
"Some Unexpected Results in Elementary Mathematics," by Ivan Niven, University of Oregon.
"N. Karmarkar's Linear Programming Algorithm," by Jim Burke, University of Oregon.

Short Presentations:

"The Cutting Edge: Boundaries of Manifolds," by Scott Metcalf, Eastern Kentucky University.
"General Discussion on an Introductory Statistics Course: What Topics Should Be Included?" by Frank Dietrich and James McKenney, Northern Kentucky University.
"A Unique Analysis Course for Secondary Teachers Getting a Master's Education," by Austin French, Georgetown College.
"A Fast Algorithm for Approximating Polynomial Roots," by Chris Christensen, Northern Kentucky University.

Short Course:

"Computer Graphics," by Gail Wells, Northern Kentucky University.

## Ohio Section

The University of Akron at Akron, Ohio was the setting for the annual spring meeting of the Ohio Section held on April 12-13, 1985. There were 146 registrants plus approximately 50 students in attendance.

Invited Addresses:

"Symbolic Computation on Modern Computers," by Paul Wang, Kent State University.
"The Embeddability of Graphs," by Phil Huneke, Ohio State University.

"Engineering Needs and the College Mathematics Core," by Isaac Greber, Case Western Reserve University.
"Three P's for Teaching Mathematics," by Jim Leitzel, Ohio State University.
"EMPT Update for Department Chairs," by Bert Waits, Ohio State University.

Contributed Papers:

"Status Report on the Revision of Teacher Certification Rules," by Alan Osborne, Ohio State University.
"Transformations Helping in Geometry," by Kenneth Cummins, Kent State University.
"Ratios of Generalized Fibonacci Sequences," by Tom Dence, Ashland College.
"Electronics for Computer Science," by Dean Brown, Youngstown State University.
"OAC Basketball Schedule Problem," by Neil Bernstein, Marietta College.
"Contest Problems from the 1985 AHME," by Dick Horwath, John Carroll University.
"Handicapped Matrices," by Russ Smucker, Muskingum College.
"Parabolic Blending Yields Piecewise Cubics for Interpolating Curves," by Cliff Long, Bowling Green State University.

Special Session:

Panel on Scientific Computing, by Tom Price, University of Akron (Moderator).
"A Computer Algorithm for the Tsirelson Space Norm," by Johnnie Baker, Kent State University.
"Providing a Complex Number Environment for MACSYMA," by Oberta A. Slotterbeck, Hiram College.
"Piecewise Linear Approximation of Implicitly Defined Manifolds," by P.H. Schmidt, University of Akron.
"Turning the Corner on Parallel Processing," by Douglas R. Smith, Goodyear Aerospace Corporation.

Student Papers:

"Computer Generation of Finite Maximal Connected Topologies," by Tom Batchik, University of Akron.
"Create Your Own Geometry," by David Cameron, Miami University.
"Counting the Infinite," by Sheldon Degenhardt, Wittenberg University.
"Accelerating Archimedes' Method for Approximating Pi," by Raymond E. Flanery, Jr., Youngstown State University.
"Interest Rate Determination: The Impact of Federal Deficits," by Timothy Fuerst, Ohio Northern University.
"Algebraic Properties of Magic Squares," by Jim Gerbic, John Carroll University.
"A First Look at C(X)," by David Goloff, University of Dayton.
"Modeling: Vented Loudspeaker Design," by Mike Grinkemeyer, Miami University.
"Motion of a Satellite," by Linda Hudson, University of Akron.
"An Implementation of an Elementary Cryptographic Technique Using Pascal," by Jeffrey N. Jones, Baldwin Wallace College.
"Column Nim," by Stephen Kurzhals, Xavier University.
"Keeping the User Friendly," by Paul Mullins, Youngstown State University.
"A Software Conversion Problem," by Mary Beth Pretorius, Wittenberg University.
"Fourier Analysis and Its Role in Resonance Technologies," by Arun Ranchod, College of Wooster.
"Conversion of ATS6/GEOS3 SST DATA to Free Air Gravity Anomalies Over the Southeast Pacific," by Terry J. Sabaka, University of Akron.
"A Special Class of Quadratic Diophantine Equations," by John Smith, John Carroll University.
"Polygonal Residue Patterns," by Karen Trageser, Miami University.
"Figurate Numbers," by Theresa Wiencek, Youngstown State University.

## Nebraska Section

Hastings College, Hastings, Nebraska was the location of the annual meeting of the Nebraska Section held on April 12-13, 1985. There were 29 registrants from 9 colleges in Nebraska and Southeastern South Dakota.

Invited Addresses:

"Statistics and the Law," by Mary Gray, American University.

Short Presentations:

* "Report on the High School Mathematics Examination," by Richard Barlow, Kearney State College.
"Triangles with Concurrent Assorted Medians, Angle Bisectors, and Altitudes," by Dale M. Mesner, University of Nebraska at Lincoln.
"How Not to Tile Rooms," by Leo G. Chouinard, University of Nebraska at Lincoln.
"Use of Spread Sheets as a Mathematical Tool," by Dale W. Behrens, Hastings College.
"An Existence Theorem for Differential Equations," by Allan Peterson, University of Nebraska at Lincoln.
"Some Good Classroom Problems," by Alexander Mehaffey, Jr., University of South Dakota.
"Continuous Functions Which Are the Composition of Retractions for Finite $T_0$-Spaces," by Mel Thornton, University of Nebraska at Lincoln.
"$dy/dx = -x^3/(x+y)$: A Visual Odyssey," by Thomas Shores, University of Nebraska at Lincoln.
"Report on TEAM Project," by Albert Zechmann, University of Nebraska at Lincoln.

Student Paper:

"Mathematical Anxiety in Arab Students Inside Israel," by Nimer Baya´a, American University.

Panel:

"Changes in the Mathematics Curricula," by Roger Olson, Dana College; Randall Heckman, Kearney State College; Dick Vogt, Nebraska Wesleyan University; David Logan, University of Nebraska at Lincoln; John Konvalina, University of Nebraska at Omaha; Wallace A. Raab, University of South Dakota.

## Missouri Section

The annual spring meeting of the Missouri Section was held April 12-13, 1985 at Central Missouri State University, Warrensburg, Missouri. There were 77 registrants.

Invited Addresses:

"Mathematics--Structure and Use; Or, What Are We Learning This For?" by Raymond Freese, St. Louis University.
"Bits and Pieces from the Classroom," by Leonard Gillman, University of Texas at Austin.

Short Presentations:

"Problem Solving and Problem Solvers," by Sherralyn Craven, Central Missouri State University.
"Applying SAS to Solve Problems in Estimating Ordinary Least Squares Regression," by Baird Brock, Central Missouri State University.
"A Two-Person Coin Flipping Problem," by Curtis Cooper and Robert E. Kennedy, Central Missouri State University.
"An Investigation of the Natural Density of $\{x: x \text{ is a factor of } f(x)\}$," by Robert E. Kennedy and Curtis Cooper, Central Missouri State University.
"How to Make Mathematics Interesting in the Classroom," by Peter G. Casazza, University of Missouri at Columbia.
"Square Free Number Cycles," by Dale Woods, Central State University (Oklahoma)/Northeast Missouri State University.
"Non-commutative Lattices," by Jonathan Leech, Missouri Western State College.
"Random Normed Structures," by Troy Hicks, University of Missouri at Rolla.

Student Presentations:

"A Mathematical Model of Human Thought," by Janet C. Tremain, University of Missouri at Columbia.
"A Graphical ´Machine´ for the Hyperbolic Functions," by David Rodriguez, Central Missouri State University.

## Southwestern Section

The annual spring meeting of the Southwestern Section was held April 12-13, 1985 jointly with the American Mathematical Society and the Sociedad Matematica Mexicana, at the University of Arizona in Tucson, Arizona. There were 120 registrants.

Invited Addresses:

"Hilbert as a Household Word," by Constance Reid.
"Some Observations on Mathematics and Mathematicians," by Ivan Niven, University of Oregon.
"Modern Real Algebra," by Gregory Brumfiel, Stanford University.
"Representable Functors among Categories of Algebras," by George M. Bergman, University of California at Berkeley.

Short Presentations:

"The Simplest Quad," by Alvin Swimmer, Arizona State University.
"Archimedes, Fibonacci, and Pascal: A Rational Search for Roots," by Robert J. Wisner, New Mexico State University.
"On the Metaphysics Which Underlies the Hypotheses Which Underlie Geometry, Part II," by Justin McCarthy, Deming, New Mexico.
"What is the Relation Between Galileo´s Work on the Free Fall and the Teaching of Mathematics?" by Alejandro Lopez-Yanez, U. Nacional Autonoma de Mexico.
"Putting the Ellipse in Perspective," by Charles G. Moore, Northern Arizona University.

Student Papers:

"Polynomials with Small Value Set Modulo p," by Javier Gomez-Calderon, University of Arizona.
"A Method for Solving Linear Differential Equations," by Ray Griesan, University of Arizona.
"Totally Orthogonal Groups All of Whose Nonlinear Characters Have the Same Degree," by Kwang Shang Wang, University of Arizona.

Now $|S: F| = |P| = p$, a prime, and thus $F$ is a maximal subfield of $S$. Also,

$$F \subseteq F_{r-1} \cap S \subseteq F_r \cap S \subseteq S.$$

Since $\alpha \notin F_{r-1} \cap S$ and $\alpha \in F_r \cap S$, we conclude that $F = F_{r-1} \cap S$ and $F_r \cap S = S$ so that $S \subseteq F_r$.

Write $E = \langle S, F_{r-1} \rangle$. Then $F_{r-1} \subseteq E \subseteq F_r$. Also, $S$ is Galois over $\mathbb{Q}$ and so natural irrationalities applies and we conclude that $|E: F_{r-1}| = |S: F| = p$ and $E$ is Galois over $F_{r-1}$. However, $F_{r-1}[\gamma_r] = F_r \subseteq \mathbb{R}$ and some power of $\gamma_r$ lies in $F_{r-1}$. Our proposition now applies and yields $p = |E: F_{r-1}| \leqslant 2$. This contradiction proves the theorem. (See Fig. 4.) ■

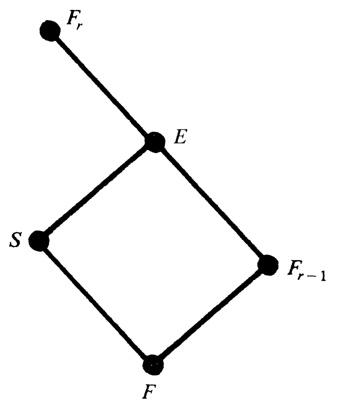



FIG. 4

We close with the observation that solvable polynomials with real roots but which are not solvable by real radicals seem to abound. For example, for any prime $p$, the polynomial

$$f(x) = x^3 - 2px + p$$

has this property. (Note that it is irreducible by the Eisenstein criterion and it has three real roots since $f(0) > 0$ and $f(1) < 0$.) It is amusing to solve this polynomial by Cardan's method to see where nonreal numbers come in.

**References**

1. E. Artin, Galois Theory, Notre Dame Math. Lectures, 1944.
2. B. L. van der Waerden, Modern Algebra (vol. 1, 2nd edition), Ungar, New York, 1953.

## A THEOREM IN COMBINATORICS AND WILSON'S THEOREM

HARAGAURI N. GUPTA
*Department of Mathematics and Statistics, University of Regina, Regina, Canada S4S 0A2*

The purpose of this note is to state and prove a theorem in combinatorics and deduce from it the famous Wilson's theorem as a special case.

THEOREM 1. *Let $B(n,r)$, $1 \leqslant r \leqslant n$, denote the number of distributions of $n$ distinct objects among $r$ ordered boxes, none remaining empty. Then,*

(i) *if $n$ is prime, then $n$ divides $B(n,r)$ for all $r \geqslant 2$,*

(ii) *if $n$ is prime, then $n$ divides $B(n-1,r) + (-1)^r$, for all $r$, $1 \leqslant r \leqslant n-1$.*

*Proof.* We shall not require the explicit formula for $B(n,r)$, but will make do with the following identities I and II, both of which can be established solely by combinatorial arguments:

I. $B(n,r) = r(B(n-1,r) + B(n-1,r-1)),\quad 2 \leqslant r \leqslant n-1.$
II. $B(n,r) = \sum_{k=1}^{n-r+1}\binom{n}{k}B(n-k,r-1),\quad 2 \leqslant r \leqslant n.$

Now, if $n$ is prime and $r \geqslant 2$, then $n-r+1 \leqslant n-1$ and each of the coefficients $\binom{n}{k}$ on the right-hand side of II is divisible by $n$. (See Remark 2 at the end of this note.) Hence $B(n,r)$ is divisible by $n$. This proves part (i) of the Theorem.

Moreover, since $n$ does not divide $r$, the identity I together with part (i) now yields:

$$(*)\qquad n \text{ divides } B(n-1,r) + B(n-1,r-1),\qquad 2 \leqslant r \leqslant n-1.$$

To prove part (ii) we apply induction on $r$. In case $r=1$, $B(n-1,1)=1$, because there is just one way of distributing $n-1$ objects in one single box. In this case we have, therefore,

$$B(n-1,1) + (-1)^1 = 1 - 1 = 0,$$

and $n$ divides 0. Assume then that $n$ divides $B(n-1,r-1) + (-1)^{r-1}$ for some $r-1$, $1 \leqslant r-1 \leqslant n-1$. By $(*)$ above.

$$B(n-1,r) + B(n-1,r-1) \equiv 0 \pmod n,\qquad 2 \leqslant r \leqslant n-1.$$

We add $(-1)^{r-1}$ to both sides of the equivalence. By the inductive hypothesis,

$$B(n-1,r-1) + (-1)^{r-1} \equiv 0 \pmod n.$$

Hence

$$B(n-1,r) \equiv (-1)^{r-1} \pmod n,$$

which amounts to:

$$B(n-1,r) + (-1)^r \equiv 0 \pmod n.$$

This completes the induction, proving part (ii) of Theorem 1.

REMARK 1. $B(n-1,n-1) = n-1!$, because the number of distributions of $n-1$ objects among $n-1$ boxes, no box being empty, is clearly $n-1!$. This then can be seen as giving a proof of Wilson's theorem.

COROLLARY (Wilson's Theorem). $(n-1)! \equiv -1 \pmod n$ *for n prime.*

REMARK 2. If $1 \leqslant k \leqslant n-1$, then $k\binom{n}{k} = n\binom{n-1}{k-1}$, as is easily established by a combinatorial argument. If $n$ is prime, then since $n \nmid k$, we have $n \mid \binom{n}{k}$.

By an argument similar to that in the paper one can establish that if $n$ is prime, then $n$ divides $\binom{n-1}{r-1} + (-1)^r$.

REMARK 3. Note that $B(n,r)/r!$ is the Stirling number of the second kind $S(n,r)$. $S(n,r)$ denotes the number of partitions of $n$ objects into $r$ cells.

## *P*-ADIC BINOMIAL COEFFICIENTS MOD *P*

ROGER C. ALPERIN
*Department of Mathematics, The University of Oklahoma, Norman, OK 73019*

In this note we provide a simple formula for evaluating the binomial coefficient

$$\binom{\alpha}{n} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!} \bmod p$$

for any $p$-adic integer $\alpha \in \mathbb{Z}_p$. As a corollary we obtain another proof of the well-known formula for the Legendre symbol $\left(\frac{2}{p}\right)$.

Let $p$ be a prime number; for a rational number $\alpha = q/r$ where the highest power of $p$ dividing $q$ is $a$, that of $r$ is $b$, we define the $p$-adic valuation, $v_p(\alpha) = a - b$. The completion of the rational numbers, $\mathbb{Q}$, with respect to the norm $|\alpha|_p = p^{-v_p(\alpha)}$ is the topological field of $p$-adic numbers, $\mathbb{Q}_p$. The elements of $\mathbb{Q}_p$ with nonnegative $p$-adic value form the ring of $p$-adic integers, $\mathbb{Z}_p$. Every $\alpha \in \mathbb{Z}_p$ has a unique $p$-adic expansion $\alpha = a_0 + a_1 p + a_2 p^2 + \cdots$ with $0 \leqslant a_i < p$, $a_i \in \mathbb{Z}$. The least nonnegative residue of $\alpha \in \mathbb{Z}_p \bmod p$ is the first $p$-adic coefficient $a_0$. Also, using the $p$-adic expansion, it is easy to see that the set $N$ of natural numbers is dense in $\mathbb{Z}_p$.

LEMMA 1. *If $\alpha \in \mathbb{Z}_p$, then $\binom{\alpha}{n} \in \mathbb{Z}_p$ for all $n \in \mathbb{N}$.*

*Proof.* The polynomial function $f(\alpha) = \binom{\alpha}{n}$ is continuous and $f(m)$ is integral valued for $m \in \mathbb{N}$. Hence $f$ maps the closure of $\mathbb{N}$, $\mathbb{Z}_p$, to $\mathbb{Z}_p$. ■

Since the value of $\binom{\alpha}{n}$ for $\alpha \in \mathbb{Z}_p$ is a $p$-adic integer, we may consider its evaluation $\bmod p$.

LEMMA 2. *If $\alpha \in \mathbb{Z}_p$, then for $n \in \mathbb{N}$*

(1) $\binom{p\alpha}{n} = 0 \bmod p$ *if $p \nmid n$,*

(2) $\binom{p\alpha}{n} = \binom{\alpha}{k} \bmod p$ *if $n = pk$.*

*Proof.* For the first part notice that it is true for $\alpha$, a natural number, by considering the coefficient of $x^n$ in $(1+x)^{p\alpha} = (1+x^p)^\alpha \bmod p$. The function $g(\alpha) = \binom{p\alpha}{n}$ is continuous in the $p$-adic variable $\alpha$ and $g(\mathbb{N}) = 0 \bmod p$; thus $g(\alpha) = 0 \bmod p$, by the density of $\mathbb{N}$ in $\mathbb{Z}_p$.

For the second part notice that by cancelling factors of $p$ and reducing $\bmod p$,

$$\begin{aligned}\binom{p\alpha}{pk} &= \frac{p\alpha}{pk}\,\frac{p\alpha-1}{pk-1}\cdots\frac{p\alpha-p}{pk-p}\cdots\frac{p\alpha-pk+1}{1}\\ &= \frac{\alpha}{k}\,\frac{\alpha-1}{k-1}\cdots\frac{\alpha-k+1}{1} = \binom{\alpha}{k} \bmod p. \blacksquare\end{aligned}$$

Now consider the power series expansion for $(1+x)^\alpha$, $\alpha \in \mathbb{Z}_p$ [**1**],

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n.$$

THEOREM. *If $\alpha \in \mathbb{Z}_p$ has the p-adic expansion $\alpha = a_0 + a_1 p + \cdots$ and $n \in \mathbb{N}$ has the p-adic expansion $n = n_0 + n_1 p + \cdots + n_s p^s$, then*

$$\binom{\alpha}{n} = \prod_{i=0}^{s} \binom{a_i}{n_i} \bmod p.$$

*Proof.* Observe that

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = (1+x)^N (1+x)^{p^{s+1}\beta}$$

where

$$\alpha = N + p^{s+1}\beta, N = a_0 + a_1 p + \cdots + a_s p^s.$$

Thus

$$\begin{aligned}(1+x)^\alpha &= \prod_{i=0}^{s} (1+x^{p^i})^{a_i} (1+x^{p^{s+1}})^\beta \bmod p\\ &= \prod_{i=0}^{s} \sum_{t=0}^{a_i} \binom{a_i}{t} x^{tp^i} \left( \sum_{k=0}^{\infty} \binom{\beta}{k} x^{kp^{s+1}} \right) \bmod p.\end{aligned}$$

Comparing coefficients of $x^n$ in the expression above we obtain the formula of the theorem.

(Notice that no terms $\binom{\beta}{k}$ enter into the coefficient of $x^n$.) ■

This theorem gives some insight into the $p$-adic expansion. It follows easily from this theorem that the $p$-adic expansion for $\alpha$ is

$$\alpha = \sum_{n=0}^{\infty}\left(\binom{\alpha}{p^n} \bmod p\right)p^n.$$

From this expression one can obtain formulas for the coefficients of the $p$-adic expansion for $\alpha+\beta$ in terms of the coefficients of $\alpha$ and $\beta$. If $\alpha = \Sigma_{i=0}^{\infty} a_i p^i$, $\beta = \Sigma_{i=0}^{\infty} b_i p^i$, then $\alpha+\beta = \sum_{i=0}^{\infty} c_i p^i$ where

$$c_n = a_n + b_n + \sum_{s=0}^{n-1}\binom{a_s+b_s}{p}\sum_{i=s+1}^{n-1}\binom{a_i+b_i}{p-1} \bmod p.$$

The details for the proof of this formula are left to the interested reader.

COROLLARY 1. *If $n = a_0 + a_1 p + \cdots + a_s p^s$ is a natural number, then*

$$\binom{2n}{n} = (-1)^n 2^{2n} \prod_{i=0}^{s}\binom{\frac{1}{2}(p-1)}{a_i} \bmod p.$$

*Proof.* From the uniqueness of $p$-adic expansions and the geometric series for $1/(1-p)$, it follows that

$$-\frac{1}{2} = \frac{p-1}{2}(1+p+p^2+\cdots)$$

is the $p$-adic expansion of $-\frac{1}{2}$. Evaluating $\binom{-\frac{1}{2}}{n}$ by means of the theorem and combining it with the formula

$$\binom{-\frac{1}{2}}{n} = (-1)^n 2^{-2n}\binom{2n}{n}$$

we obtain the stated result. ■

According to Euler's Theorem, for an odd prime $p$,

$$\left(\frac{2}{p}\right) = 2^{\frac{1}{2}(p-1)} \bmod p.$$

Using the next corollary we can obtain the evaluation of $\left(\frac{2}{p}\right)$.

COROLLARY 2. For an odd prime $p$, $2^{\frac{1}{2}(p-1)} = (-1)^{(p^2-1)/8} \bmod \mathrm{p}$.

*Proof.* If $p = 1 \bmod 4$, let $n = (p-1)/4$ and apply the previous corollary. If $p = -1 \bmod 4$, let $n = (p+1)/4$ and apply the previous corollary. ■

**Reference**

1. N. Koblitz, $P$-adic Numbers, $P$-adic Analysis and Zeta-Functions. Springer-Verlag, 1977.

---

**ANSWER TO PHOTO ON PAGE 559**

The picture shows Grace Chisholm Young (1868–1944) who with her husband William Henry Young (1862–1943) produced fundamental research in general topology, measure theory, and functional analysis—and also six children, who include Professor L. C. Young of the University of Wisconsin. The picture was taken before her marriage, in fact probably in the mid 1890s when she took her doctorate at Göttingen, the first in the new program.

$$= \prod_{n\geqslant 1} (1 + x^{2n-1})(1 + x^{2n})(1 - x^{2n})$$
$$= \prod_{n\geqslant 1} (1 + x^{2n-1})(1 - x^{4n})$$
$$= \prod_{n\geqslant 1} (1 + x^{4n-3})(1 + x^{4n-1})(1 - x^{4n}),$$

and we have

$$\prod_{n\geqslant 1}\left(\frac{1-x^n}{1+x^n}\right)^2 = 1 - 4\sum_{n\geqslant 1}\left(\frac{x^{4n-3}}{1+x^{4n-3}} - \frac{x^{4n-1}}{1+x^{4n-1}}\right). \tag{4}$$

Now,

$$\prod_{n\geqslant 1}\left(\frac{1-x^n}{1+x^n}\right) = \prod_{n\geqslant 1}\frac{(1-x^{2n-1})(1-x^{2n})}{(1+x^n)}$$
$$= \prod_{n\geqslant 1}(1-x^{2n-1})(1-x^n)$$
$$= \prod_{n\geqslant 1}(1-x^{2n-1})(1-x^{2n-1})(1-x^{2n})$$
$$= \sum_{-\infty}^{\infty}(-1)^n x^{n^2},$$

so (4) is

$$\left(\sum_{-\infty}^{\infty}(-1)^n x^{n^2}\right)^2 = 1 - 4\sum_{n\geqslant 1}\left(\frac{x^{4n-3}}{1+x^{4n-3}} - \frac{x^{4n-1}}{1+x^{4n-1}}\right). \tag{5}$$

Put $-x$ for $x$, and we obtain

$$\left(\sum_{-\infty}^{\infty} x^{n^2}\right)^2 = 1 + 4\sum_{n\geqslant 1}\left(\frac{x^{4n-3}}{1-x^{4n-3}} - \frac{x^{4n-1}}{1-x^{4n-1}}\right), \tag{6}$$

from which Theorem 1 follows immediately [**2**].

## References

1. John A. Ewell, A simple proof of Fermat's two-square theorem, this MONTHLY, 90 (1983) 635–637.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960, p. 258.

# A SIMPLE STRUCTURE THEOREM FOR TWO-GENERATOR, ONE-RELATION GROUPS

DONALD P. MINASSIAN
*Department of Mathematical Sciences, Butler University, Indianapolis, IN 46208*

A group $G$, with normal subgroup $N$, is said to be the *extension* of $N$ by the factor group $G/N$. That is, $G$ admits a two-step normal series: $G \supseteq N \supseteq \{e\}$. If $N$ and $G/N$ have readily grasped structure, then valuable insights can be gained about the structure of $G$. In this note we shall consider the important class of groups given by $G_n = \{a,b|ba = ab^n\}$, where $n$ is any fixed natural number exceeding 1. Specialists well know [**1**] that each such group is an extension of a subgroup $N_n$ of the additive rational numbers by the additive integers $G/N_n$. However, a proof for nonspecialists appears lacking in the literature. This note fills that gap. (The skeletal proof in [**1**] deals with $G_2$. The author knows of no reference for $G_n$, $n > 2$.)

We assume that $G_n$ has two "independent elements" or generators $a$ and $b$ of infinite order,

that $G_n$ consists of all elements of form $a^x b^y$ where $x$ and $y$ are integers, and that $a$ moves past $b$ by employing the given relation $ba = ab^n$. We need two more interrelated definitions:

DEFINITION 1. For any fixed natural number $n$, an *n-adic rational number* is a rational number of form $\pm n^k$ where $k$ is any integer.

DEFINITION 2. The set $H_n$ of all finite sums (including zero) of $n$-adic rational numbers is clearly a subgroup of the additive rational numbers. We call $H_n$ the *n-adic rational group*.

For example, $-1/16 = -2^{-4}$ is a 2-adic (or dyadic) rational number, and $H_2 = \{\Sigma_k \pm 2^k,$ integer $k\}$ is the dyadic rational group.

The reader may have rightly surmised that $N_n$ of the first paragraph turns out to be $H_n$. Note that $H_n$ contains the integers, since $\Sigma_{i=1}^n(1/n) = 1$. Speaking loosely, $H_n$ may be deemed the "second simplest subgroup" within the additive rational numbers after groups isomorphic to the integers. (We say "second simplest" since if $H_n$ contained a minimal power of $n$, say $n^m$, then $H_n$ would be generated by $n^m$ and hence isomorphic to the integers.) Henceforth, we drop the subscript $n$ from $G_n$, $N_n$, and $H_n$.

We first show that $N$ of the first paragraph is precisely that subgroup of $G$ consisting of all finite products of $n$-adic powers of $b$, i.e., $N = \{\Pi_w b^{n^w}\}$, for integers $w$, is the $n$-adic rational group $H$. We then show that $N$ is normal in $G$ and that $G/N$ is an infinite cyclic group ("the additive integers"). We first need a Lemma and its corollary.

LEMMA. *For any integer $w$, $b^{n^w} = a^{-w}ba^w$. (Strictly speaking, $b^{n^w}$ is not yet defined for negative $w$ but will be below.)*

*Proof.* By definition of $G$ we see that $b^n = a^{-1}ba$. Now

$$b^{n^2} = (b^n)^n = (a^{-1}ba)^n = a^{-1}b^n a = a^{-1}(a^{-1}ba)a = a^{-2}ba^2;$$

further,

$$b^{n^3} = (b^{n^2})^n = (a^{-2}ba^2)^n = a^{-2}b^n a^2 = a^{-2}(a^{-1}ba)a^2 = a^{-3}ba^3;$$

and generally,

$$b^{n^k} = a^{-k}ba^k, \quad \text{for } k = 0,1,2,\dots . \tag{1}$$

Also, for $k = 0,1,2,\dots,$

$$\begin{aligned}(a^k ba^{-k})^{n^k} &= a^k(b^{n^k})a^{-k}\\ &= a^k(a^{-k}ba^k)a^{-k}, \text{ by } (1)\\ &= b,\end{aligned}$$

and thus $a^k ba^{-k}$ is an $n^k$th root of $b$—call it $b^{n^{-k}}$. (We might ask, "Is the $n^k$th root of $b$ unique?" The answers are (a) it doesn't affect our argument either way, and (b) "yes" (see [**1**]; alternatively, I shall supply a proof upon request for general $G_n$)). That is, we *define*

$$b^{n^{-k}} \equiv a^k ba^{-k}, \quad \text{for } k = 0,1,2,\dots . \tag{2}$$

Together, (1) and (2) prove the Lemma.

COROLLARY. *If $x$ and $y$ are integers with $x \geqslant y$, then $b^{n^x} = (b^{n^y})^{n^{x-y}}$, where $b^{n^x}$ and $b^{n^y}$ are given by the Lemma.*

*Proof.* Simply calculate

$$(b^{n^y})^{n^{x-y}} = (a^{-y}ba^y)^{n^{x-y}} = a^{-y}b^{n^{x-y}}a^y = a^{-y}(a^{y-x}ba^{x-y})a^y = a^{-x}ba^x = b^{n^x},$$

where all equalities except the second come from the Lemma.

Three propositions will now complete our task.

PROPOSITION 1. *The subgroup $N$ of $G$ generated by $b^{n^w}$ where $w$ ranges over all integers—that is, the set $N$ of all finite products of the form $b^{n^{w_1}}b^{n^{w_2}} \cdots b^{n^{w_m}}$—is an $n$-adic rational group*, $N \cong H$.

*Proof.* Let $H$ be the $n$-adic rational group as defined in Definition 2. Let each $\alpha_i$ be an $n$-adic rational number; that is, $\alpha_i$ has the form $\pm n^{w_i}$, for some integer $w_i$. Note that as usual $b^{-n^w}$ means the group inverse of $b^{n^w}$. The map $v$ from $H$ to $N$ given by

$$\left(\sum_{i=1}^{n} \alpha_i\right) v \equiv \prod_{i=1}^{n} b^{\alpha_i}$$

looks suspiciously like an isomorphism. We show that it is and the proof follows.

(i) The map $v$ is well-defined. Suppose $\Sigma_{i=1}^{p}\alpha_i = \Sigma_{j=1}^{q}\beta_j$, where each $\alpha_i$ and $\beta_j$ are $n$-adic rational numbers: $\alpha_i = \pm n^{w_i}$ and $\beta_j = \pm n^{r_j}$ for integers $w_i$ and $r_j$. Now clearly one (or more) of the $\alpha_i$ and $\beta_j$ are *smallest* in absolute value; i.e., the absolute value of at least one of them is $n^m$, where $m$ is the smallest of all the exponents $w_i$ and $r_j$. So the Corollary above implies that each $b^{\alpha_i} = (b^{n^m})^{\alpha_i/n^m}$, since $\alpha_i = n^{m_i}$, where $m_i \geqslant m$; similarly for each $b^{\beta_j}$. Thus the following two relations are equivalent:

$$\text{(I)} \qquad \prod_{i=1}^{p} (b^{\alpha_i}) = \prod_{j=1}^{q} (b^{\beta_j}),$$

and

$$\text{(II)} \qquad \prod_{i=1}^{p} \left[(b^{n^m})^{\alpha_i/n^m}\right] = \prod_{j=1}^{q} \left[(b^{n^m})^{\beta_j/n^m}\right].$$

Since *each* $\alpha_i/n^m$ and $\beta_j/n^m$ is integral, (II) is clearly equivalent (by elementary rules of integer exponents in groups) to

$$\text{(III)} \qquad (b^{n^m})^{\sum_{i=1}^{p} \alpha_i/n^m} = (b^{n^m})^{\sum_{j=1}^{q} \beta_j/n^m}.$$

"Well-defined" follows since $\Sigma\alpha_i = \Sigma\beta_j$ by assumption!

(ii) The map $v$ is a homomorphism onto $N$. This comes from the definitions of $H$, $N$ and the well-defined map $v$.

(iii) The map $v$ is one-to-one. If not, $b^{\Sigma\alpha_i} = e$ (the group identity) for some nonzero $\Sigma\alpha_i$. Let $w$ be the minimum of the exponents $w_i$ for all $\alpha_i = \pm n^{w_i}$. Thus $(\Sigma_i\alpha_i)/n^w$ is integral. If $w \geqslant 0$, making $n^w$ integral, then by the rules of integer exponents in groups,

$$\text{(3)} \qquad (b^{n^w})^{\Sigma\alpha_i/n^w} = b^{\Sigma\alpha_i} = e, \text{ where } \Sigma\alpha_i/n^w \neq 0 \text{ (since } \Sigma\alpha_i \neq 0 \text{ by assumption)}.$$

Thus a nonzero integer power of $b$ is the identity. This is impossible since $b$ has infinite order. If $w < 0$, recall by (2) that $b^{n^w}$ raised to the power $n^{-w}$ is $b$. Now raise both sides of (3) to the positive integer power $n^{-w}$. By properties of integer exponents in groups, $n^{-w}$ commutes past the integer exponent given by the first $\Sigma$ in (3) to produce

$$b^{\Sigma\alpha_i/n^w} = e^{n^{-w}} = e,$$

and again some nonzero integer power of $b$ is the identity.

Because the map $v$ is well-defined, $b^\alpha$ is well-defined for a sum $\alpha$ of $n$-adic rational numbers $\alpha_i = \pm n^w$, regardless of how $\alpha$ is expressed as such a sum $\Sigma \pm n^w$ for integers $w$. That is, $b^\alpha = \prod b^{\pm n^w}$ is well-defined. (See [**2**]. p. 159, for a simple example.) The following two proposi-

tions complete our task:

PROPOSITION 2. *The subgroup $N$ consisting of all $n$-adic rational powers of $b$ is normal in $G$.*

*Proof.* For any integers $w$ and $i$ we have

$$b^{n^{w+i}} = a^{-(w+i)}ba^{w+i} = a^{-i}(a^{-w}ba^{w})a^{i} = a^{-i}b^{n^{w}}a^{i},$$

by using the Lemma in the first and third equalities. Thus (reversing the steps), $a^{-i}b^{n^w}a^i = b^{n^{w+i}} \in N$ for any integer $w$. Further, taking inverses of both sides of this last equality, $a^{-i}b^{-n^w}a^i \in N$. Thus, for any $\alpha = \Sigma_w \pm n^w$ we have

$$a^{-i}b^{\alpha}a^{i} = \prod_{w}(a^{-i}b^{\pm n^{w}}a) \in N, \text{ for any integer } i.$$

Since $b$ is itself an $n$-adic rational power of $b(b^1 = b^{n^0} \in N)$, $N$ is trivially invariant under conjugation by $b$ and $b^{-1}$. Thus the definition of $G$ as being generated by $a$ and $b$ and the fact that $N$ is invariant under conjugation by any integer power of $a$ or $b$ yields the desired normality of $N$ in $G$.

PROPOSITION 3. *The factor group $G/N$ is infinite and cyclic.*

*Proof.* (i) Cyclic. By the definitions of $G$ and $N$, $G/N$ is generated by the coset $aN$.

(ii) Infinite. If not, then $(aN)^k = N$ and so $a^k = b^\alpha$ for some nonzero integer $k$ and some $\alpha = \Sigma_w n^w$ for integers $w$. Now there exists a large enough nonzero integer $m$ (a power of $n$) such that $m\alpha$ is an integer, $s$. Raising both sides of $a^k = b^\alpha$ to the $m$th power, we have $a^r = b^s$, for integers $r$ and $s$, where $r = mk \neq 0$ since both $m$ and $k \neq 0$. Without loss of generality, take $r$ positive. Thus $a^r \equiv b^s$ would commute with $b$, $ba^r = a^rb$, whereas the definition of $G$ shows that $ba^r = a^rb^t$, where $t$ is a natural number greater than 1, $\neq a^rb$, since $b$ has infinite order as noted in the second paragraph of this note. (It is at this point that the argument would break down if we let $n = 1$ in the definition of $G$. However, the abelian group $G = \{a, b | ba = ab\}$ is simple enough in its own right. Clearly, $N$ consists of all "1-adic rational powers" of $b$—i.e., all integral powers of $b$—and thus is isomorphic to the additive integers. Trivially, $N$ is normal in $G$ and the subgroup $G/N$ is cyclic. Also, $G/N$ would still be infinite, i.e., we could not have $a^r = b^s$, but the argument would involve a further discussion of "free generators", which we have chosen to avoid here.)

Thus the group $G$ is indeed "metabelian" (two-step abelian), i.e., $G$ has a normal series $G \supseteq N \supseteq \{e\}$, where $N$ and $G/N$ are both abelian. Furthermore, $N$ and $G/N$ are among the least complicated of all infinite groups.

## References


1. L. Fuchs and B. Sasiada, Note on orderable groups, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. (1964) 13–17.
2. Neal H. McCoy, Introduction to Modern Algebra, 3rd ed., Allyn and Bacon, Boston, 1975.


## CORRECTION TO "NOTE ON THE CAYLEY-HAMILTON THEOREM"


MARVIN JAY GREENBERG
*Department of Mathematics, University of California, Santa Cruz, CA 95064*


Formula (4), p. 194, this MONTHLY, March 1984, need not hold even when $B$ commutes with $A$, because if $Q(t) = \operatorname{adj}(tI - A)$, $Q(B)$ will not in general equal $\operatorname{adj}(B - A)$. All we can say if $B$ commutes with $A$ is that

$$Q(B)(B - A) = \chi_A(B) = (B - A)Q(B).$$

My thanks to John R. Silvester for this correction.

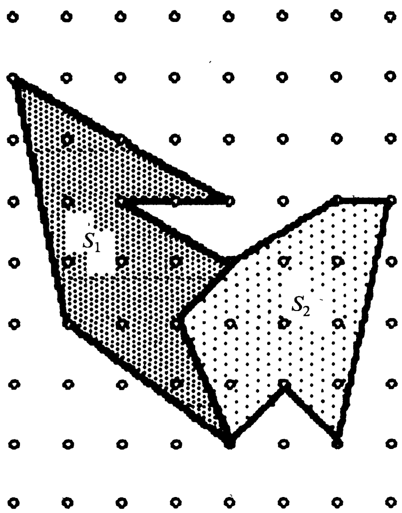
# PICK'S THEOREM REVISITED

DALE E. VARBERG
*Department of Mathematics, Hamline University, St. Paul, MN 55104*

Pick's theorem is one of the gems of elementary mathematics, this because the most innocent sounding hypotheses imply a very surprising conclusion. Yet the statement of the theorem can be understood by a fifth grader. Call a polygon simple if its boundary is a simple closed curve and call it a lattice polygon if the coordinates of its vertices are integers. Pick's theorem asserts that the area of a simple lattice polygon $S$ is given by

$$A(S) = i + \frac{1}{2}b - 1 = v - \frac{1}{2}b - 1$$

where $i$, $b$, and $v$ are, respectively, the number of interior lattice points, the number of boundary lattice points, and the total number of lattice points of $S$ (Fig. 1).




$$A(S) = i + \frac{b}{2} - 1$$

$$A(S_1) = 7 + \tfrac{9}{2} - 1 = \tfrac{21}{2}$$

$$A(S_2) = 7 + \tfrac{7}{2} - 1 = \tfrac{19}{2}$$

$$A(S_1 \cup S_2) = 15 + \tfrac{12}{2} - 1 = 20$$

FIG. 1

We are aware of several proofs of Pick's theorem [**1**], [**2**], [**4**], [**5**], [**6**]. Each of them uses an elegant trick or two but finally rests on the stubborn fact that a primitive triangle (one whose only lattice points are its vertices) has area 1/2. We call this a stubborn fact because its derivation seems to be the major hurdle in all the standard proofs.

Our proof of Pick's theorem is direct, intuitive, and requires nothing more sophisticated than the almost obvious and easily proved fact (proof by induction will work) that a lattice polygon can be decomposed as a union of lattice triangles.

To begin, associate with each lattice point $P_k$ of $S$ a weight $w_k = \theta_k/2\pi$, where $\theta_k$ measures the "visibility" angle with which $P_k$ can see into $S$. Thus $w_k = 1$ at an interior lattice point, $w_k = 1/2$ at a boundary lattice point that is not a vertex, and $w_k = 1/4$ at a right-angled corner point. Think of $w_k$ as measuring the contribution that $P_k$ makes to the area of $S$ (see Fig. 2). Let

$$W(S) = \sum_{P_k \in S} w_k$$

and, as before, let $A(S)$ be the area of $S$.

LEMMA. $W(S) = A(S)$.

*Proof*. We note first that $W$ is additive; that is, if $S = S_1 \cup S_2$ as in Fig. 1, then $W(S) =$

$W(S_1) + W(S_2)$. This is a consequence of the fact that the visibility angles in $S_1$ and $S_2$ at a common lattice point add together to give the visibility angle in $S$ at that point.

Next consider successively (i) a lattice rectangle with sides parallel to the lattice (Fig. 2), (ii) a lattice right triangle with legs parallel to the lattice (Fig. 3), and (iii) an arbitrary lattice triangle. The latter can be surrounded by right triangles of type (ii) to form a rectangle of type (i) (Fig. 4 shows a typical example).

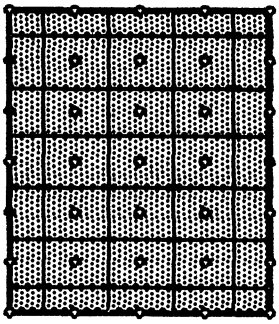


FIG. 2

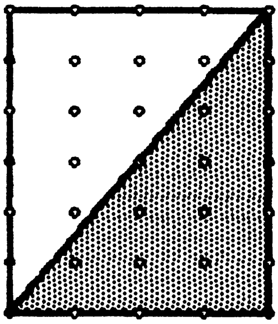


FIG. 3

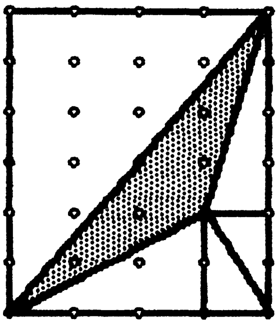


FIG. 4

That $W(S) = A(S)$ in Case (i) is obvious and Case (ii) follows immediately from Case (i) upon division by 2. We obtain the result in Case (iii) by using the shared additivity of $W$ and $A$. It remains only to decompose an arbitrary lattice polygon as a union of lattice triangles and then to apply the additivity of $W$.

We remark that the simpleness of $S$ is not used in the proof of the lemma. ■

PICK'S THEOREM. *For a simple lattice polygon,*

$$A(S) = i + \frac{1}{2}b - 1 = v - \frac{1}{2}b - 1.$$

*Proof.* A simple polygon with $c$ interior vertex angles has angle sum $(c-2)\pi$. For example, a triangle has angle sum $\pi$, a quadrilateral $2\pi$, and so on. It follows that the sum of all the visibility angles at points $P_k$ along the boundary of $S$ is $(b-2)\pi$. Thus if $I$ and $B$ denote the interior and boundary of $S$, then

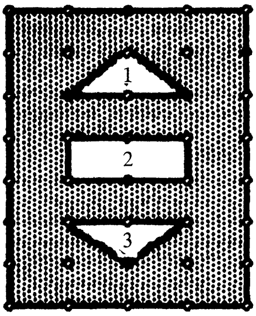
$$\begin{aligned} A(S) = W(S) &= \sum_{P_k \in I} w_k + \sum_{P_k \in B} w_k \\ &= i + \frac{(b-2)\pi}{2\pi} \\ &= i + \frac{1}{2}b - 1. \blacksquare \end{aligned}$$

**Generalizations.** Consider next a general lattice polygon $S$, one that may have holes or intersect itself (see Figs. 5, 6, and 7). We require only that $S$ can be expressed as the union of finitely many simple lattice polygons. Pick's theorem as originally stated does not apply to such figures but a simple variant involving the Euler characteristic $\chi$ does. The correct formula is

$$A(S) = v - \frac{1}{2}e_b - \chi.$$

Here, $e_b$ is the number of boundary edges of $S$. Note that for a simple lattice polygon, $e_b = b$ and $\chi = 1$.

The situation illustrated in Fig. 5 is easily handled with no additional apparatus. Let $S$ be a simple lattice polygon containing $m$ separated holes which are themselves simple lattice polygons. Recall that the Euler characteristic for such a figure is $1 - m$. Let $b_0, b_1, b_2, \ldots, b_m$ denote the number of lattice points along the outer boundary of $S$ and the boundaries of the $m$ holes of $S$,

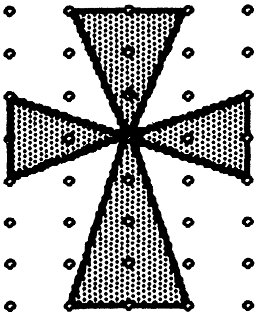



$v = 40 \qquad e_b = 36$

$\chi = -2$

$A(S) = 40 - \frac{1}{2}(36) + 2$

FIG. 5

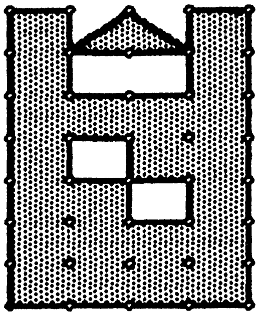


$v = 20 \qquad e_b = 16$

$\chi = 1$

$A(S) = 20 - \frac{1}{2}(16) - 1$

FIG. 6



$v = 40 \qquad e_b = 38$

$\chi = -2$

$A(S) = 40 - \frac{1}{2}(38) + 2$

FIG. 7

respectively. Then, using the fact that the visibility angles of the $k$th hole sum to $(b_k + 2)\pi$, we obtain

$$\begin{aligned} A(S) &= i + \frac{1}{2\pi}\sum_{P_k \in B} \theta_k \\ &= i + \frac{(b_0 - 2)}{2} + \frac{(b_1 + 2)}{2} + \cdots + \frac{(b_m + 2)}{2} \\ &= i + \frac{1}{2}(b_0 + b_1 + \cdots + b_m) + (m - 1) \\ &= i + \frac{1}{2}b - \chi = v - \frac{1}{2}b - \chi. \end{aligned}$$

This demonstration breaks down for the polygons of Figs. 6 and 7. For them, we are forced to follow the standard procedure of decomposing $S$ into primitive triangles each of area 1/2, together with a simple combinatorial argument. We state the general result as a theorem.

EXTENDED PICK'S THEOREM. *Let $S$ be a lattice polygon (simple or not). Then*

$$A(S) = v - \frac{1}{2}e_b - \chi,$$

*where $v$ is the total number of lattice points in $S$, $e_b$ is the number of edges on the boundary of $S$, and $\chi$ is the Euler characteristic of $S$.*

*Proof.* We suppose that $S$ has been decomposed into primitive triangles (see [4] for a proof that this can be done) and let $v$, $e$, and $f$ denote the number of vertices, edges, and faces in this decomposition. Each triangle has 3 edges and each edge is shared by 2 triangles except for those edges on the boundary of $S$. Thus, $3f = 2e - e_b$ and so

$$f = -e_b + 2e - 2f = 2v - e_b - 2(v - e + f) = 2v - e_b - 2\chi.$$

We conclude that

$$A(S) = \frac{1}{2}f = v - \frac{1}{2}e_b - \chi. \blacksquare$$

We have recently learned of an even more general form of Pick's theorem; it appears in reference [3].

**References**

**1.** W. W. Funkenbusch, From Euler's formula to Pick's formula using an edge theorem, this MONTHLY, 81 (1974) 647–648.

**2.** R. W. Gaskell, M. S. Klamkin, and P. Watson, Triangulations and Pick's theorem, Math. Mag., 49 (1976) 35–37.

**3.** H. Hadwiger and J. M. Wills, Neuere Studien über Gitterpolygone, J. Math., 280 (1975) 61–69.

**4.** R. Honsberger, Ingenuity in Mathematics, New Mathematical Library, vol. 23, Mathematical Association of America, Washington, DC, 1970, 27–31.

**5.** Andy C. F. Liu, Lattice points and Pick's theorem, Math. Mag., 52 (1979), 232–235.

**6.** I. Niven and H. S. Zuckerman, Lattice points and polygonal area, this MONTHLY, 74 (1967) 1195–1200.

## UNBOUNDED SEQUENCES OF EULER-DEDEKIND MEANS

CHARLES R. WALL
*Trident Technical College, Charleston, SC 29411*

Euler's totient function, the number of integers not exceeding $n$ and prime to it, is well known to be given by

$$\phi(n) = n\Pi(1 - 1/p),$$

where the product is taken over the distinct primes that divide $n$. Less well known is the closely related Dedekind function:

$$\psi(n) = n\Pi(1 + 1/p).$$

Define $g(n)$ to be the mean of these:

$$g(n) = (\phi(n) + \psi(n))/2.$$

It is easy to see that $g(n)$ is always an integer and that $g(n) = n$ whenever $n$ is a prime power. So, if we iterate $g$, the resulting sequence may become constant:

$$g(12) = 14,\ g(14) = 15,\ g(15) = 16,\ g(16) = 16, \dots .$$

Richard Guy [1] asked if such sequences can increase indefinitely; we show that the answer is "yes."

Let $x_{n+1} = g(x_n)$ with $x_1 = 1488 = 2^4 3 \cdot 31$. Then

$$x_2 = 2^4 3 \cdot 37, \quad x_3 = 2^6 3 \cdot 11, \quad x_4 = 2^6 41, \quad x_5 = 2^5 83, \quad x_6 = 2^4 167,$$
$$x_7 = 2^3 5 \cdot 67, \quad \text{and} \quad x_8 = 2^5 3 \cdot 31 = 2x_1,$$

so that for $n \geqslant 1$, we have $x_{n+7} = 2x_n$ and the sequence $\{x_n\}$ is unbounded.

The two smallest initial values of such unbounded sequences are $x_1 = 45$ and $x_1 = 50$. Each of these gives $x_3 = 56$, $x_{28} = 1488$, so that for $n \geqslant 28$, we again have $x_{n+7} = 2x_n$.

**Reference**

**1.** Richard K. Guy, Unsolved Problems in Number Theory, Springer-Verlag, New York, 1981.

---

**155.** **MISCELLANEA**

What is the next element of the sequence that begins with F4E, S9, SE5EN, ... ?

—Donald E. Knuth

See p. 595.

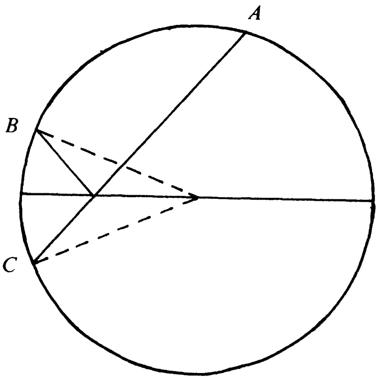



FIG. 1

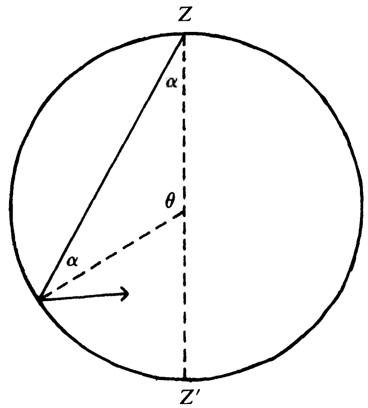



FIG. 2

Also solved by K. L. Bernstein, J. Dou (Spain), V. Grinberg, L. R. King, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), A. Markovich (Yugoslavia), W. W. Meyer, A. Müller (Switzerland), B. Olk, M. Pachter (South Africa), F. B. Strauss, B. Vidakovic (Yugoslavia).

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1986. The solver's full post-office address should be on each sheet.*

6501. *Proposed by I. J. Schoenberg, Madison, Wisconsin.*

Suppose that we have an identity

$$e^z \equiv \left(\sum_{n=0}^{+\infty} a_n z^n\right)\left(\sum_{n=0}^{+\infty} b_n z^n\right) \quad \text{for} \quad |z| < r, (r > 0), \tag{1}$$

where $a_n \geqslant 0$ and $b_n \geqslant 0$ for all $n$, and neither of the factors in (1) reduces to a constant. Show that the two factors in (1) are necessarily of the form

$$e^{az+c} \quad \text{and} \quad e^{bz-c}, \quad \text{with} \quad a > 0, b > 0, a + b = 1. \tag{2}$$

6502. *Proposed by Klaus Schürger, University of Bonn, Federal Republic of Germany.*

Each of the following conditions is clearly necessary for the convergence of a sequence $\{S_n\}$:

(A) $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} S_k$ exists; and (B) $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} S_{m2^k}$ exists for every integer $m \geqslant 1$.

If $\{S_n\}$ is a real sequence which is bounded above and $\{nS_n\}$ is nondecreasing, is either condition sufficient for the convergence of $\{S_n\}$?

---

## ANSWER TO QUESTION ON PAGE 587

E1GHT. The rule is to replace the letters that are used as Roman numerals by their Arabic equivalents.

## THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

### MORE—AND MOORE—POWER SERIES WITHOUT TAYLOR'S THEOREM

I. E. LEONARD
*Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115*

JAMES DUEMMEL
*Department of Mathematics, Western Washington University, Bellingham, WA 98225*

Inspired in part by the article "Power series without Taylor's theorem" by Wells Johnson in the June-July 1984 issue of the MONTHLY, we present here an extremely simple method for obtaining the power series expansions for the functions $\sin x$, $\cos x$, $e^{-x}$, $e^{x}$. The approach we take uses no "sophisticated" concepts such as integration by parts, reduction formulas, or recursion relations; it uses only the monotonicity of the definite integral, and the basic inequalities for the functions in question. We make no claim of originality: the first author learned of the technique from Professor R. A. Moore, at Carnegie-Mellon University, while the second author encountered it in a calculus text (see [**2**], p. 78).

Starting with the inequality, $\cos t \leqslant 1$, and integrating repeatedly over the interval $[0, x]$, we obtain inequalities involving the successive Taylor polynomials for sine and cosine. For example, four integrations yield, for $x \geqslant 0$,

$$x - \frac{x^3}{3!} \leqslant \sin x \leqslant x$$

and

$$1 - \frac{x^2}{2!} \leqslant \cos x \leqslant 1 - \frac{x^2}{2!} + \frac{x^4}{4!}.$$

Students are quickly convinced that $\sin x$ and $\cos x$ are trapped between successive partial sums of their usual Taylor series expansions. The error estimates

$$\left|\cos x - \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!}\right| \leqslant \frac{x^{2n+2}}{(2n+2)!}$$

and

$$\left|\sin x - \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!}\right| \leqslant \frac{x^{2n+3}}{(2n+3)!}$$

are evident, and a little more argument will convince even the most skeptical students that the series actually converge to $\sin x$ and $\cos x$.

As soon as the derivative of the exponential function has been covered, this approach will also yield the Taylor series for $e^{-x}$ from the inequality $e^{-t} \leqslant 1$, for $t \geqslant 0$. This series can be used to estimate $e^{-1}$ and hence $e$. In particular, the estimate $(24/9) < e < 3$ is easily obtained in this way. This could be used below for students who find $3^b$ more acceptable than $e^b$.

The power series expansion for $e^x$, $x \geqslant 0$, can be obtained from the inequality $1 \leqslant e^t \leqslant e^b$, for $0 \leqslant t \leqslant b$ ($b$ fixed), by integrating repeatedly, and we obtain, after rearranging,

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{n+1}}{(n+1)!} \leqslant e^x \leqslant 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{n+1}}{(n+1)!} e^b$$

for $0 \leqslant x \leqslant b$. This inequality gives the error estimate

$$0 \leqslant e^x - \sum_{k=0}^{n+1} \frac{x^k}{k!} \leqslant \frac{x^{n+1}}{(n+1)!}(e^b - 1)$$

for $0 \leqslant x \leqslant b$. Now a simple argument showing that $\lim_{n\to\infty}(x^n/n!) = 0$, for fixed $x$, yields the power series expansions for $\sin x$, $\cos x$, $e^{-x}$, $e^x$, valid for $x \geqslant 0$. Note that taking $x = b = 1$ in the above inequality gives

$$\frac{1}{(n+1)!} \leqslant e - \sum_{k=0}^{n} \frac{1}{k!} \leqslant \frac{e}{(n+1)!}.$$

### References

1. W. Johnson, Power series without Taylor's theorem, this MONTHLY, 91 (1984) 367–369.
2. J. F. Randolph, Calculus, Macmillan, New York, 1952.

## AN ELEMENTARY PROOF OF A THEOREM IN CALCULUS

DONALD E. RICHMOND
*500 Fulton St., #201, Palo Alto, CA 94301*

A fundamental theorem states that if $f'(x) = 0$ at every point of an interval $[a, b]$, then $f(x)$ is constant on $[a, b]$. Most proofs use the Mean Value Theorem, but several proofs have been given that are independent of this theorem (Bers [**1**], Halperin [**3**], Powderly [**4**]). The present note proves our fundamental theorem in an elementary way.

Let $f'(x) \equiv 0$ on $[a, b]$. If $f(x)$ is not constant on $[a, b]$ as the theorem states, then for some $u < v$ on $[a, b]$, $f(v) \neq f(u)$. Hence the chord joining $(u, f(u))$ and $(v, f(v))$ has a slope different from zero. That is

(1) $$f(v) - f(u) = C(v - u), \quad \text{or} \quad \Delta f = C\Delta x, \quad \text{with } C \neq 0.$$

Assume that $C > 0$. Bisect $[u, v]$ at $w$. If

$$f(v) - f(w) < C(v - w)$$

and

$$f(w) - f(u) < C(w - u),$$

then

$$f(v) - f(u) < C(v - u),$$

which contradicts (1) with $C > 0$. Hence on at least one of the intervals $[u, w]$ and $[w, v]$, $\Delta f/\Delta x \geqslant C$. Call this interval $[u_1, v_1]$. By repeated bisection one obtains a nested sequence of intervals $[u_n, v_n]$ over each of which $\Delta f/\Delta x \geqslant C$. $[u_n, v_n]$ converges to some $x$ in $[u, v]$. Hence if $C > 0$, $\Delta f/\Delta x$ cannot approach $f'(x) = 0$ as required.

The possibility $C < 0$ is excluded by reversing the inequalities. Hence $C = 0$, contrary to the assumption that $f(v) \neq f(u)$.

The same argument shows that if $f'(x) \geqslant 0$ on $[a, b]$, then $f(v) \geqslant f(u)$, and if $f'(x) \leqslant 0$ on $[a, b]$, then $f(v) \leqslant f(u)$.

The argument may be generalized to show that if $m \leqslant f'(x) \leqslant M$ on $[a, b]$, then

(2) $$m(v - u) \leqslant f(v) - f(u) \leqslant M(v - u).$$

It is sufficient to replace $C > 0$ by $C > M$ and $C < 0$ by $C < m$. (2) has been called the "weak" form of the Mean Value Theorem [**2**].

### References

1. L. Bers, On avoiding the mean value theorem, this MONTHLY, 74 (1967) 583.
2. R. P. Boas, Who needs those mean-value theorems, anyway?, Two-Year College Mathematics Journal, 12 (1981) 178–181.
3. I. Halperin, A fundamental theorem of the calculus, this MONTHLY, 61 (1954) 122–123.
4. M. Powderly, A simple proof of a basic theorem of the calculus, this MONTHLY, 70 (1963) 544.

# PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1986. Please place the solver's name andd mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3105. *Proposed by Calin P. Popescu, student, Bucharest, Romania, and the editors.*

Let $X$ be a compact space (not necessarily $T_1$) and $f\colon X \to X$ a function with the property that $f(A) \subseteq A$ for every closed subset $A$ of $X$. Show that there is a singleton set $P$ such that $P \subseteq \overline{f(P)}$ and a closed set $Q$ such that $\overline{f(Q)} = Q$.

E 3106. *Proposed by Donald E. Knuth, Stanford University.*

Let $S(n)$ be the set of all positive integers $k$ such that the fractional part of $n/k$ is $1/2$ or more. For example,

$$S(17) = \{2,3,6,9,10,11,18,19,20,\dots,34\}.$$

Prove that

$$\sum_{k\in S(n)} \phi(k) = n^2,$$

where $\phi$ is Euler's totient function.

E 3107. *Proposed by Ira Gessel, Brandeis University.*

Show that if $m$ and $n$ are integers, then $\dfrac{m!(2m+2n)!}{(2m)!n!(m+n)!}$ is an integer.

E 3108. *Proposed by L. Cseh and I. Merényi (students), Cluj, Romania.*

Let $x, y, z$ be real numbers, $k_1, k_2, k_3 \in (0, 1/2)$ and $k_1 + k_2 + k_3 = 1$. Prove that

$$k_1 k_2 k_3 (x + y + z)^2 \geqslant xyk_3(1 - 2k_3) + yzk_1(1 - 2k_1) + zxk_2(1 - 2k_2).$$

When does equality hold?

E 3109. *Proposed by Desmond MacHale and Mícheál Ó Searcóid, University College, Cork, Ireland.*

Let $R$ be an associative ring such that $(xy)^n = yx$, for all $x, y \in R$, for some fixed natural number $n$. Must $R$ be commutative?

E 3110. *Proposed by Dennis Spellman, Sacred Heart University, Bridgeport, CT.*

Let $p$ be an odd prime. Show that if $(x_0, y_0)$ is a pair of positive rational numbers on the curve $x^p + y^p = 1$, then at $(x_0, y_0)$ both $dx/ds$ and $dy/ds$ are quadratic irrational numbers. Here $s$ denotes arc length.

## SOLUTIONS OF ELEMENTARY PROBLEMS

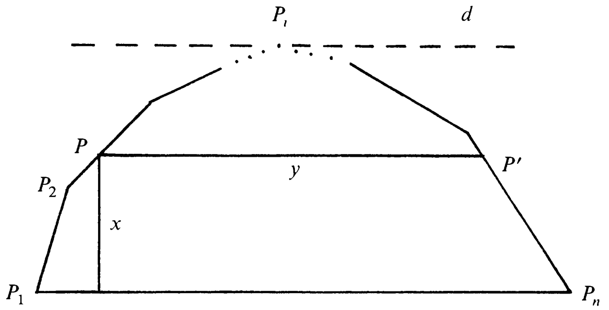
### A Square in a Polyhedron

E 2740 [1978, 765]. *Proposed by Victor Pambuccian, Bucharest, Romania.*

Show that if $P$ is a convex polyhedron, one can find a square all of whose vertices are on some three faces of $P$, as well as a square whose vertices are on four different faces of $P$.

*Partial solution by the proposer.* We first prove the following

LEMMA. *Let $P_1 P_2 \ldots P_n (n \geqslant 3)$ be a convex polygon, such that $\sphericalangle P_2 P_1 P_n < 90°$ and $\sphericalangle P_1 P_n P_{n-1} < 90°$ and $P_1 P_n$ is not parallel to any $P_i P_{i+1}$ $(i = 1, \ldots, n-1)$. Then one can find a square all of whose vertices lie on exactly three sides of the polygon.*




*Proof.* Let $d$ be the support line parallel to $P_1 P_n$ and $P_t$ the contact point, $P_t = d \cap P_1 P_2 \ldots P_n$. For $P \in P_1 P_2 \ldots P_n$ let $x = d(P, P_1 P_n)$ (the distance from $P$ to $P_1 P_n$) and $y = |PP'|$, where $P' \in P_1 P_2 \ldots P_n$ and $PP' \| P_1 P_n$. Put $d(P_t, P_1 P_n) = x_0$. Then

$$f: [0, x_0] \to R \qquad f(x) = y$$

is a continuous function. So $g$: $[0, x_0] \to R, g(x) = f(x) - x$ is also continuous and $g(0) =$

$|P_1P_n| > 0$ and $g(x_0) = -x_0 < 0$. Since $g$ has the intermediate value property, $g(x_1) = 0$ for some $x_1 \in (0, x_0)$. Q.E.D.

THEOREM. *For every convex polyhedron there is a square, all of whose vertices lie on some three distinct faces of the polyhedron.*

*Proof.* Let $V$ be any vertex of the polyhedron and $\alpha_1, \alpha_2, \ldots, \alpha_n$ be all the faces which meet in $V$ (i.e., $V \in \alpha_i, i = 1, 2, \ldots, n$). Let $d_{ij} = \alpha_i \cap \alpha_j$ be the lines of intersection of the planes containing faces $\alpha_i$ and $\alpha_j$. Take two points $A$ and $B \in \alpha_1$ on the two edges that belong to $\alpha_1$ and pass through $V$, such that $AB \nparallel d_{ij}$ for $i, j = 2, \ldots, n$, and $\sphericalangle VAB < 90°$, $\sphericalangle VBA < 90°$. (The possibility of such a choice is obvious.) Consider now all the planes that pass through $AB$. When these planes are sufficiently near to $\alpha_1 = (VAB)$, their intersection with the polyhedron will be a convex polygon with the angles at $A$ and $B$ acute and having no side parallel to $AB$. Applying the previous lemma to such a polygonal section we obtain the desired result.

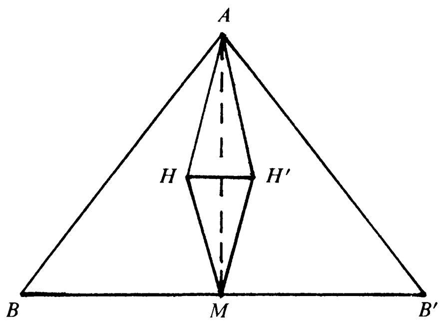
### A Symmetric Dissection of an Isosceles Triangle

E 2918 [1981, 763]. *Proposed by Jordi Dou, Barcelona, Spain.*

Show that an isosceles triangle can be dissected symmetrically around the principal median into seven acute isosceles triangles except when the vertex angle $A$ is 90°, 120°, or when $135° \leqslant A \leqslant 144°$.

*Solution by the proposer.* Let $ABB'$ ($B = B'$) be the triangle. If $A < 90°$, we obtain a partition into 4 equal triangles. With another partition into 4, of one of the two central triangles, one obtains a partition into 7.

Let $A \geqslant 90°$. Let $H$ be the unique interior vertex (on the axis) of the partition: if there are 2 interior symmetric vertices $H$ and $H'$ not on the axis, we sketch the triangles $AHH'$ and $MHH'$, and by symmetry will have an even number of triangles. If $H$ and $H'$ are on the axis, sketch $AHP$, $AHP'$, and $MH'M'$; then the pentagon $BMH'HP$ cannot possibly be partitioned into two triangles. (See Fig. 1.)



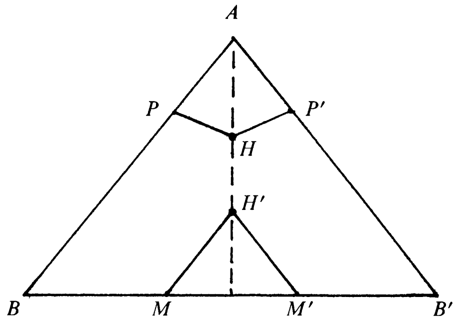




FIG. 1

Let $P, P'$ be the (necessary) vertices on $AB, AB'$; $M, M'$ those situated on $BB'$. The "principal" vertices of the triangles $BPM$, $BP'M'$ are $B$ and $B'$ (because $B = B' < 45°$). That of $HMM'$ is $H$ (by symmetry). For the triangles $AHP$ and $PHM$ (or $AHP'$ and $P'HM'$), we have to consider the 9 pairs of principal vertices. (See Fig. 2.)

PRINCIPAL VERTICES

$\theta, \gamma$: Then $\omega = (\pi + 2A)/4 > 90°$ is impossible.

$\theta, \delta$: Then $\delta = 90°$, which is impossible.

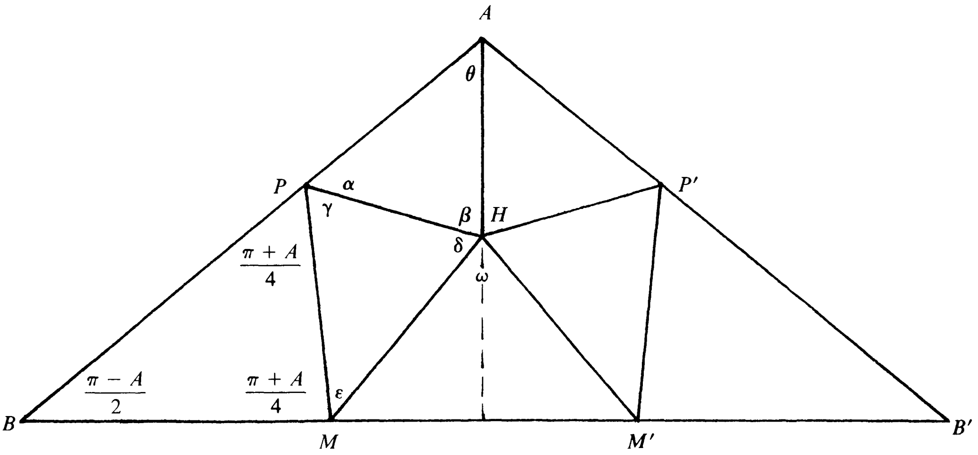



FIG. 2

$\theta, \epsilon$: Then $\epsilon = 90°$, which is impossible.

$\alpha, \gamma$: Then $\omega = (3\pi - A)/4 > 90°$ is impossible.

$\beta, \gamma$: Then $\beta = \pi - A < 90°$, and $\omega = (5A - \pi)/4 < 90°$ implies $A < 108°$. Therefore $90° < A < 108°$.

$\beta, \delta$: Then $\beta = \pi - A < 90°$ and $\delta = (3A - \pi)/2 < 90°$ implies $A < 120°$. Therefore $90° < A < 120°$.

$\beta, \epsilon$: Then $\beta = \pi - A < 90°$ and $\omega = (7A - 3\pi)/2 < 90°$ implies $A < (102.857\ldots)°$. Therefore $90° < A < (102.857\ldots)°$.

$\alpha, \delta$: Then $\delta = (3\pi - 3A)/2 < 90°$ implies $A > 120°$, and $\omega = 2A - \pi < 90°$ implies $A < 135°$. Therefore $120° < A < 135°$.

$\alpha, \epsilon$: Then $\omega = (5\pi - 5A)/2 < 90°$ implies $A > 144°$. Therefore $A > 144°$.

Also partially solved by W. Janous (Austria).

**An Even Combinatorial Sum**

E 2932 [1982, 212; 1982, 498]. *Proposed by Henry E. Fettis, Mountain View, CA.*

For $n > 1$, set

$$S_n(b) = \sum_{k=0}^{[(n+b)/2]} (-1)^k \frac{(n - 2k + b)^{n-1}}{k!(n-k)!}$$

Prove that, if $b < n, S_n(b) = S_n(-b)$.

*Solution by Kenneth L. Bernstein, MITRE Corporation, Bedford, MA.*

$$S_n(b) = \frac{1}{n!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \times (\text{polynomial in } k \text{ of degree } n-1)$$

$$- \sum_{k=[(n+b)/2]+1}^{n} (-1)^k \frac{(n-2k+b)^{n-1}}{k!(n-k)!}.$$

According to Euler's formula (a proof of which appears in this MONTHLY, 85 (1978) 450–467) the first sum is zero. Substituting $n - s$ for $k$ leaves

$$S_n(b) = \sum_{s=0}^{n-1-[(n+b)/2]} (-1)^s \frac{(n-2s-b)^{n-1}}{s!(n-s)!}.$$

When $n+b$ is odd the upper limit in the above sum is $[(n-b)/2]$. When $n+b$ is even the upper limit in the above sum is $[(n-b)/2]-1$, but this limit may be extended to $[(n-b)/2]$ since, for this value, the summand vanishes. Thus:

$$S_n(b) = \sum_{k=0}^{[(n-b)/2]} (-1)^k \frac{(n-2k-b)^{n-1}}{k!(n-k)!}$$
$$= S_n(-b) \text{ Q.E.D.}$$

*Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* According to Feller, *An introduction to probability theory and its applications*, vol. 2, first ed., p. 27,

$$u_n(x) = 2^{-n}/(n-1)! \sum_{k=0}^{[x/2]} (-1)^k \binom{n}{k} (x-2k)^{n-1} \tag{1}$$

is the density of the probability distribution of the sum of $n$ independent random variables having uniform distributions on $(0,2)$. It follows that $u_n$ is symmetric around $x=n$.

Hence we have

$$S_n(b) = \frac{2^n}{n!}(n-1)!u_n(n+b)$$
$$= \frac{2^n}{n}u_n(n+b)$$

which is symmetric around $b=0$.

Replacing 2 in (1) by an arbitrary $a>0$ yields an obvious generalization.

Also solved by P. S. Bruckman, L. Carlitz, C. S. Karruppan Chetty (India), I. Gessel, O. G. Ruehr, University of South Alabama Problem Group, C. Wildhagen (The Netherlands), and the proposer.

Carlitz noted that $n!S_n(2k-n) = 2^{n-1}A_{n-1,k}$, where $A_{n-1,k}$ is an Eulerian number (see, for instance, page 243 of Louis Comtet's *Advanced Combinatorics*, Reidel Publishing, 1974).

**Light Ray in a Semicircle**

E 2950 [1982, 424]. *Proposed by Ko-Wei Lih, Academia Sinica, Taiwan.*

The inner side of a semicircle (including diameter) is a mirror. A light ray emitting from the zenith makes an angle $\alpha$ with the vertical line, $0 \leqslant \alpha \leqslant \pi/2$. Characterize $\alpha$ such that the light ray will hit the zenith after finitely many reflections. (The context can be enlarged to a hemispherical mirror, including horizontal base.)

*Solution by the proposer.* Observe that if we have a circular mirror, as in Fig. 1, the light ray $AC$ will hit $B$ when it is reflected by a horizontal mirror through the center. Since $B$ and $C$ are equidistant from the horizon, instead of following the light ray in the semicircular mirror, we may follow the same light ray in the circular mirror as in Fig. 2. When it hits a point on the lower circumference, the original light ray hits the symmetric point on the upper half. When it hits a point on the upper circumference, so does the original light ray. Therefore we must characterize the angle $\alpha$ in the circular mirror such that the light ray hits $Z$ or $Z'$ after finitely many reflections. This occurs if and only if $\theta/\pi = m/n$, where $m$ and $n$ are integers with $0 < m \leqslant n$ and $(m,n)=1$. Thus $\alpha/(\pi/2) = (n-m)/n$, provided we make the convention that if a light ray hits a corner, it reflects back in the same direction. If we do not allow the light ray to hit the corners, then $n$ must be odd.

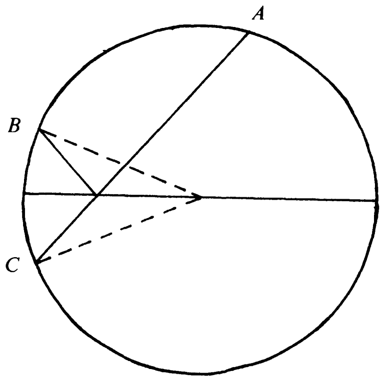



FIG. 1

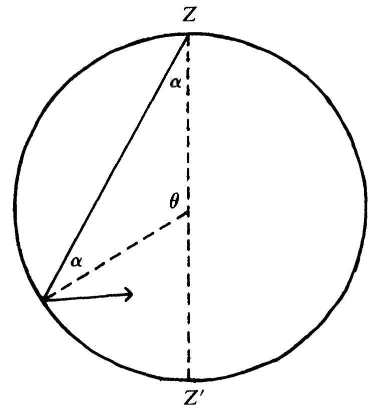



FIG. 2

Also solved by K. L. Bernstein, J. Dou (Spain), V. Grinberg, L. R. King, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), A. Markovich (Yugoslavia), W. W. Meyer, A. Müller (Switzerland), B. Olk, M. Pachter (South Africa), F. B. Strauss, B. Vidakovic (Yugoslavia).

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by January 31, 1986. The solver's full post-office address should be on each sheet.*

6501. *Proposed by I. J. Schoenberg, Madison, Wisconsin.*

Suppose that we have an identity

$$e^z \equiv \left(\sum_{n=0}^{+\infty} a_n z^n\right)\left(\sum_{n=0}^{+\infty} b_n z^n\right) \quad \text{for} \quad |z| < r, (r > 0), \tag{1}$$

where $a_n \geqslant 0$ and $b_n \geqslant 0$ for all $n$, and neither of the factors in (1) reduces to a constant. Show that the two factors in (1) are necessarily of the form

$$e^{az+c} \quad \text{and} \quad e^{bz-c}, \quad \text{with} \quad a > 0, b > 0, a + b = 1. \tag{2}$$

6502. *Proposed by Klaus Schürger, University of Bonn, Federal Republic of Germany.*

Each of the following conditions is clearly necessary for the convergence of a sequence $\{S_n\}$:

(A) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} S_k$ exists; and (B) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} S_{m2^k}$ exists for every integer $m \geqslant 1$.

If $\{S_n\}$ is a real sequence which is bounded above and $\{nS_n\}$ is nondecreasing, is either condition sufficient for the convergence of $\{S_n\}$?

---

## ANSWER TO QUESTION ON PAGE 587

E1GHT. The rule is to replace the letters that are used as Roman numerals by their Arabic equivalents.

## SOLUTIONS OF ADVANCED PROBLEMS

### Sets with Dense Projections

6421 [1983, 134]. *Proposed by J. Beck and J. Pach, Hungarian Academy of Sciences, and F. Galvin, University of Kansas.*

Let $E$ be a set of points in the plane with the property that every closed disc of radius 1 includes at least one element of $E$. Prove that there exists a straight line $L$ such that the orthogonal projection of $E$ onto $L$ is everywhere dense in $L$.

*Correction to solution.* M. J. Pelling has pointed out that the statement of the theorem used in the published solution [1984, 589] is incorrect. The unit circle

$$E = \{(x, y) : x^2 + y^2 = 1\}$$

evidently satisfies the hypothesis of the theorem since

$$\{x/y : (x, y) \in E, y \neq 0\} = \mathbb{R},$$

but, for every $\alpha \in \mathbb{R}$, $\{x + \alpha y : (x, y) \in E\}$ is bounded and so is not dense in $\mathbb{R}$. The defect can be eliminated by replacing "$\{x/y : (x, y) \in E, y \neq 0\}$ is dense in $\mathbb{R}$" in the hypothesis of the theorem by "for every $n \in \mathbb{Z}^+$, $\{x/y : (x, y) \in E, |y| \geqslant n\}$ is dense in $\mathbb{R}$". The set $E$ in the corollary satisfies the amended condition.

### A Determinant That Rarely Vanishes

6455 [1984, 205]. *Proposed by L. E. Mattics, University of South Alabama.*

Let $a_{i1} = (-1)^{i+1}$, $a_{ij} = i^{j-1}$ for $1 \leqslant i \leqslant p-1$, $2 \leqslant j \leqslant p-1$, where $p$ is an odd prime. Show that $\det(a_{ij}) \equiv 0 \pmod p$ if and only if $\sum_{k=1}^{p-1}(-1)^{k+1}k^{-1} \equiv 0 \pmod p$. Is there a prime $p$ for which $\det(a_{ij}) \equiv 0 \pmod p$?

*Solution by the editors.* Unfortunately a misprint occurred in the statement of the problem. It should have read $a_{ij} = i^j$ instead of $a_{ij} = i^{j-1}$. The correct matrix $(a_{ij})$ is then

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 2^2 & 2^3 & \cdots & 2^{p-1} \\ 1 & 3^2 & 3^3 & \cdots & 3^{p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & (p-1)^2 & (p-1)^3 & \cdots & (p-1)^{p-1} \end{pmatrix}.$$

For $k = 1, 2, \ldots, p-1$, divide the $k$th row by $k$ and then add all the rows to the final row to obtain the following matrix $\pmod p$

$$\begin{pmatrix} 1 & 1 & 1 & & 1 \\ -2^{-1} & 2 & 2^2 & & 2^{p-2} \\ -3^{-1} & 3 & 3^2 & & 3^{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ -(p-2)^{-1} & (p-2) & (p-2)^2 & \cdots & (p-2)^{p-2} \\ S & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where $S = \sum_{k=1}^{p-1}(-1)^{k+1}k^{-1}$. The determinant of this matrix is $1!2!3! \cdots (p-2)!S$ and consequently $\det(a_{ij}) \equiv 0 \pmod p$ if and only if $S \equiv 0 \pmod p$. To date the only primes for which the final congruence is known to hold are $p = 1093$ [W. Meissner (1913)] and $p = 3511$ [N. G. W. H. Beerger (1922)].

Also solved by the proposer. Lajos Takács solved the printed problem with the first congruence replaced by (mod $p^2$). Various versions of the problem were solved by S. V. Kanetkar, Pierre Lalonde (Canada), O. P. Lossers (The Netherlands) and Robert E. Shafer.

**The Digamma and a Related Function**

6456 [1984, 205]. *Proposed by Robert E. Shafer, Berkeley, CA.*

For $x > 0$, define

$$\psi(x) = \lim_{n \to \infty} \log\left(n + x - \frac{1}{2}\right) - \sum_{k=0}^{n} \frac{1}{x+k}$$

and

$$\psi_2(x) = \lim_{n \to \infty} \frac{1}{2}\log^2\left(n + x - \frac{1}{2}\right) - \sum_{k=0}^{n} \frac{\log(x+k)}{x+k}.$$

(a) Show that, for all $x > 0$,

$$\psi_2(x) < \frac{1}{2}\psi^2(x) - \frac{1/24}{(x-1/2)^2 + 3/10}.$$

(b) Prove that

$$\sum_{n=1}^{\infty} (-1)^n \frac{\log n}{n} = \left(\gamma - \frac{1}{2}\log 2\right)\log 2,$$

where $\gamma$ is Euler's constant.

*Solution to part* (a) *by the proposer.* Note that $\psi(x)$ is the digamma function $\Gamma'(x)/\Gamma(x)$. By the mean value theorem we have, for $k \geqslant 1$, $x > 0$,

$$0 < \log(x+k) - \log(x+k-1) - \frac{1}{x+k} \leqslant \frac{1}{x+k-1} - \frac{1}{x+k},$$

and hence

$$0 < \log(x+n) - \log x - \sum_{k=1}^{n} \frac{1}{x+k} \leqslant \frac{1}{x} - \frac{1}{x+n}.$$

Letting $n \to \infty$, we obtain

(1) $$0 \leqslant \log x - \psi(x) \leqslant \frac{1}{x} \quad \text{for } x > 0.$$

Likewise we can show that

(2) $$0 \leqslant \frac{1}{2}\log^2 x - \psi_2(x) \leqslant \frac{\log x}{x} \quad \text{for } x \geqslant e - 1.$$

In order to prove (a) we shall first show that, for $x > 0$,

(3) $$\begin{aligned}\frac{\log x}{x} &= \psi_2(x+1) - \psi_2(x) \\ &> \frac{1}{2}\left(\psi^2(x+1) - \psi^2(x)\right) - \frac{1/24}{(x+1/2)^2 + 3/10} + \frac{1/24}{(x-1/2)^2 + 3/10}.\end{aligned}$$

Since $\psi(x+1) - \psi(x) = 1/x$, the right-hand side of inequality (3) equals

$$\frac{1}{x}\left(\psi(x) + \frac{1}{2x}\right) + \frac{x/12}{x^4 + x^2/10 + 121/400}.$$

Consequently (3) is equivalent to

$$\psi(x) < \log x - \frac{1}{2x} - \frac{x^2/12}{x^4 + x^2/10 + 121/400}. \tag{4}$$

We shall in fact prove the tighter inequality

$$\psi(x) \leqslant \log x - \frac{1}{2x} - \frac{1/12}{x^2 + 1/10}. \tag{5}$$

To prove (5) we shall first show that

$$\begin{aligned} \psi(x+1) - \psi(x) &= \frac{1}{x} \\ &> \log\frac{x+1}{x} - \frac{1}{2(x+1)} + \frac{1}{2x} - \frac{1/12}{(x+1)^2 + 1/10} + \frac{1/12}{x^2 + 1/10}, \end{aligned} \tag{6}$$

which is equivalent to

$$\log\frac{x+1}{x} < \frac{1}{2}\left(\frac{1}{x} + \frac{1}{x+1}\right) - \frac{1}{6}\,\frac{x+1/2}{(x+1/2)^4 - (3/10)(x+1/2)^2 + 49/400}. \tag{7}$$

This will follow from the tighter inequality

$$\log\frac{x+1}{x} < \frac{x+1/2}{(x+1/2)^2 - 1/4} - \frac{1/6}{(x+1/2)^3} - \frac{1/20}{(x+1/2)^5}. \tag{8}$$

Expanding in terms of $x + 1/2$, the left-hand side of (8) is

$$\sum_{n=0}^{\infty} \frac{1}{2^{2n}(2n+1)(x+1/2)^{2n+1}},$$

while the right-hand side is

$$\frac{1}{x+1/2} + \frac{1}{2^2\cdot 3(x+1/2)^3} + \frac{1}{2^4\cdot 5(x+1/2)^5} + \sum_{n=0}^{\infty}\frac{1}{2^{2n}(x+1/2)^{2n+1}}$$

and (8) follows. Thus (7) and (6) hold. From (6) we get

$$\begin{aligned} \sum_{n=0}^{N-1} (\psi(x+n+1) - \psi(x+n)) &= \psi(x+N) - \psi(x) \\ &> \log\frac{x+N}{x} - \frac{1}{2(x+N)} + \frac{1}{2x} - \frac{1/12}{(x+N)^2 + 1/10} + \frac{1/12}{x^2 + 1/10}, \end{aligned} \tag{9}$$

which is equivalent to

$$\begin{aligned} \psi(x) < \psi(x+N) - \log(x+N) &+ \frac{1}{x(x+N)} + \frac{1/12}{(x+N)^2 + 1/10} \\ &+ \log x - \frac{1}{2x} - \frac{1/12}{x^2 + 1/10}. \end{aligned} \tag{10}$$

By (1), $\psi(x+N) - \log(x+N) \to 0$ as $N \to \infty$, and so (5) follows from (10). Thus (3) is established. It follows from (3) that

$$\sum_{n=0}^{N-1} (\psi_2(x+n+1) - \psi_2(x+n)) = \psi_2(x+N) - \psi_2(x) \tag{11}$$

$$> \frac{1}{2}\left(\psi^2(x+N)-\psi^2(x)\right)-\frac{1/24}{(x+N-1/2)^2+3/10}+\frac{1/24}{(x-1/2)^2+3/10}$$

and so

$$\psi_2(x)<\psi_2(x+N)-\frac{1}{2}\psi^2(x+N)+\frac{1/24}{(x+N-1/2)^2+3/10}+\frac{1}{2}\psi^2(x)-\frac{1/24}{(x-1/2)^2+3/10}. \tag{12}$$

However, from (1) and (2), $\psi_2(x+N)-\frac{1}{2}\psi^2(x+N)\to 0$ as $N\to\infty$, and (a) follows. Strictly we can only deduce (a) with $\leqslant$ in place of $<$ from (12), but a more careful analysis shows that (5) in fact implies (a) as it stands.

A number of readers pointed out that part (b) is well known. See this MONTHLY, 87 (1980) 498 for references.

**Representations of Residue Classes**

6457 [1984, 259]. *Proposed by R. L. McFarland, Yellow Springs, OH.*

Let $p$ be an odd prime. Show that the $2^{(p-1)/2}$ numbers of the form $\pm 1\pm 2\pm\cdots\pm\frac{1}{2}(p-1)$ represent each nonzero residue class modulo $p$ the same number of times. Determine this common number of representations and show that it differs by one from the number of representations of the zero residue class.

*Solution by James Propp, University of California at Berkeley.* Let

$$\zeta=e^{2\pi i/p}\quad\text{and}\quad S=\left(\zeta+\zeta^{-1}\right)\left(\zeta^2+\zeta^{-2}\right)\cdots\left(\zeta^{(p-1)/2}+\zeta^{-(p-1)/2}\right).$$

Then

$$S=a_0+a_1\zeta+\cdots+a_{p-1}\zeta^{p-1},$$

where $a_k$ is the number of ways $k$ can be represented (mod $p$) by $\pm 1\pm 2\pm\cdots\pm\frac{1}{2}(p-1)$. Since $S$ is also equal to

$$\left(\zeta^{(p+1)/2}+\zeta^{-(p+1)/2}\right)\cdots\left(\zeta^{p-1}+\zeta^{-(p-1)}\right),$$

we obtain

$$S^2=\prod_{k=1}^{p-1}\left(\zeta^k+\zeta^{-k}\right)=\left(\prod_{k=1}^{p-1}\zeta^k\right)\left(\prod_{k=1}^{p-1}\left(1+\zeta^{-2k}\right)\right)$$

$$=\left(\zeta^{(p-1)p/2}\right)\left(\prod_{j=1}^{p-1}\left(1+\zeta^j\right)\right)=(1)\frac{(1)^p+(1)^p}{(1)+(1)}=1$$

(using the identity $(x+y)(x+y\zeta)(x+y\zeta^2)\cdots(x+y\zeta^{p-1})=x^p+y^p$). Hence $S=\pm 1$. Since $1+x+x^2+\cdots+x^{p-1}$ is the minimal polynomial of $\zeta$ over the rationals, it follows that

$$a_0-S=a_1=a_2=\cdots=a_{p-1}.$$

Further, since $\zeta+\zeta^2+\cdots+\zeta^{p-1}=-1$ and

$$a_0+(p-1)a_1=a_0+a_1+\cdots+a_{p-1}=2^{(p-1)/2},$$

we have

$$S=a_0-a_1=\left(2^{(p-1)/2}-(p-1)a_1\right)-a_1=2^{(p-1)/2}-pa_1$$

$$\equiv 2^{(p-1)/2}\bmod p,$$

so that $S = \left(\frac{2}{p}\right)$. Thus

$$a_1 = \frac{1}{p}\left(2^{(p-1)/2} - \left(\frac{2}{p}\right)\right) \quad \text{and} \quad a_0 = a_1 + \left(\frac{2}{p}\right).$$

Also solved by Paul S. Bruckman, Barry A. Cipra, L. E. Clarke (England), Stephen M. Gagola, Jr., Ray Hill (England), S. V. Kanetkar, Pierre Lalonde (Canada), David Leep, O. P. Lossers (The Netherlands), L. E. Mattics, J. G. Mauldon, R. W. K. Odoni (England), R. G. E. Pinch (Scotland), Lajos Takâcs, L. Van Hamme (Belgium), G. Velissarios (Greece), D. H. Wiedemann (Canada) and the proposer. Gerald Myerson pointed out that the problem appears as Theorem 1 of his paper, "A combinatorial problem in finite fields, I", *Pacific Journal of Mathematics*, 82 (1979), 179–187. The above solution is a bit simpler than his proof.

---

## REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Complex Variables with Applications*. By A. David Wunsch. Addison-Wesley, Reading, MA, 1983. viii + 439 pp.

GEORGE PIRANIAN
*Department of Mathematics, The University of Michigan, Ann Arbor, Michigan 48109*

Without inspecting the book, I committed myself to writing a review. When the book arrived, a mixture of incompetence and blood oozed from my pores. The MAA had beguiled me with its Editorial Guidelines for Reviewers. I had been careless about accepting a blind date, and I deserved the consequences.

The Guidelines virtually forbid chapter-by-chapter summaries as well as criticisms of format or style, and they clearly disapprove of reports on classroom testing. "Other than this, almost anything goes." Would the telling of Old-Testament stories be acceptable, and how near to it can we come in the review of a complex-variables text for undergraduates?

Somewhere in this essay we must smite the rock and hope for two or three paragraphs about the book; but first we shall praise the Guidelines. We praise them because they promote instability and because they convey the message that some of the MONTHLY's curators prefer a prophetic to a priestly orientation. I take the distinction from Kenneth E. Boulding's essay *Religious Perspectives in Economics* (pages 179 to 197 in *Beyond Economics*, *Essays on Society*, *Religion and Ethics*; Ann Arbor Paperbacks, The University of Michigan Press, 1968, 1970).

The prophet descends from the mountain or strides into the city from the desert. He may even arrive by ship or by whale. Whatever his apparel, his manners, and his station in life, he denounces well-established practices (such as sin, for example) and demands reform. His message may have minimal substantive merit; but it can claim a basic virtue: it promotes the examination of customs, of legal codes, of relations between components of society, perhaps even of the sacred conventions that separate good from evil. Occasional reassessments are essential for survival, because without them our society does not adapt to environmental changes promptly enough to remain viable. We need prophets, even as we need poets, actors, clowns, and lutenists.

The priests dedicate themselves to the perpetuation of the faith, possibly also to what we call its refinement or its elaboration depending on whether the development fills us with joy or disgust. Without priests, the available supply of prophets would be woefully inadequate for the maintenance of reasonable cultural values; but when priests become too abundant and powerful, the prophet can no longer get a hearing.

In our book reviews, frivolity would be out of place; our daily communication suffers from a smothering component of fluff contributed by newspapers, radio stations, the commercial world, local and cosmic governments, and ourselves. Agreed, then, that the MONTHLY's reviews should be solid; but if they are also stolid, we will not read them, and their publication will degenerate into a ceremony whose omission most of us would never notice. Therefore we should support the Editors' campaign for vitality, even if our concept of vitality is so diffuse that we are unable to define it.

Now we must smite the rock, and we must do it without worrying that the rock may splinter the rod. Had I known that David Wunsch is a Professor of Electrical Engineering, I would not have accepted the assignment. A chatty essay about the book's subject should make clear why the description of certain physical phenomena is most effective if we formulate it in terms of harmonic functions. Also, it should point out that every harmonic function of two real variables $x$ and $y$ is the real part of an analytic function of the complex variable $z = x + iy$. Readers of the MONTHLY are sufficiently aware of the second point, and nobody has ever accused me of being an expert on the first.

If I were to write a book about complex analysis, I would not apologize for using the letter $i$ instead of $j$; but with or without such an apology, my book would never succeed with a class of undergraduates in engineering. Watching Wunsch lead the youngsters by their hands, we marvel at the powerful flux of his affection. The author is able to share the students' collective identity. Together with a sound knowledge of the subject and the ability to formulate intelligible statements, the capacity for identification with the students forms the basic requirement for successful pedagogy.

To needle the Editors and to honor the Faithofourfathers, let us record that Chapters 6, 7, and 8 carry the titles *Residues and their use in integration*, *Laplace transforms and stability of systems*, and *Conformal mapping and some of its applications*. Section 7.3 is entitled *Principle of the argument and introduction of the Nyquist stability criterion*. Chapter 8 ends with examples illustrating the use of the Schwarz-Christoffel formula, and it has an appendix under the title *The stream function and capacitance*.

This review is not what the Editors requested; I hope they can nevertheless live with it. While we wait for perfection, we must tolerate each other's inadequacies. As to the Old-Testament tales, we never got close enough to them to use even one word of Yiddish; the more's the pity, but as the Good Book says, there is a time for everything. This includes a time for celebrating cultural affiliations and preferences and a time for pushing such sentiments into the background. Also, there is a time for treating complex analysis as if it were the North wall of the Eiger, a challenge to those who are in shape for a demanding athletic performance, and a time for presenting the subject to those who, if they are to be true to themselves, will always regard it as a tool for the solution of practical problems. Wunsch succeeds admirably.

---

*Classics of Mathematics*. By Ronald Calinger (ed.). Moore Publishing Company, Oak Park, IL, 1982. xxv + 742 pp. $18.95.

HAROLD M. EDWARDS
*New York University, Courant Institute of Mathematical Sciences, New York, NY 10012*

A mathematician who decides to look into great mathematical writing of the past faces real difficulties. He may need to acquire new terminology or even a new way of formulating mathematical ideas. Background study will no doubt be necessary to gain some notion of the mathematical environment in which his text was written. If he is studying a range of topics or periods, he is likely to have to consult many different source books. And, alas, he is likely to find many of the texts written in a language he doesn't know.

Ronald Calinger's volume does much to alleviate these difficulties. His anthology contains

translations of writers as diverse as Ptolemy and Hilbert, Bhaskara and Boole. It has 742 densely printed pages, crammed with some of the most important mathematical works of all time. There are also section introductions providing historical context for the works, and biographies of many of the writers. And, of course, it is all in English. The list price is $18.95. If there is the remotest chance you will ever want to look into the history of mathematics, you should order a copy of this bargain book.

Surely the publisher intends the book to be used as a textbook for courses in the history of mathematics, and it will be useful for this purpose as well. It is pretty heavy going, however, for any but the most dedicated and well-educated students, and there aren't many of those.

You might have imagined that the texts in this anthology are photographically reproduced from other books and that this accounts for the low price. Not at all. It has been entirely set from scratch in a two-column-per-page format, so that, as I indicated above, there is a great deal on a page. I am not happy with the type face that was used, but the proof-reading appears to have been done extremely carefully and I find practically no misprints.

It would be foolish to quibble with the selections. No two people who cared about a field could be expected to agree on such a vast number of individual choices. I disagree with some choices, but I admire the imaginativeness of others—for example, the use of a passage from Gårding which quotes Riemann on the subject of the partial differential equations of physics and then comments briefly on modern views on the subject. I feel that in very many cases (Archimedes on the quadrature of the parabola, Euler on sums of two squares) passages are cut too short to be comprehensible. But there would be no other way to get so much into the book, and these snippets may be enough to give the reader the flavor of the work and send him to complete sources if he likes the taste.

Perhaps it will give an idea of the virtues and shortcomings of the book if I examine one selection in detail. One of the selections of greatest interest to me personally is #109, Kummer's first introduction of his theory of "ideal complex numbers". It occupies 6 pages of the book, and Calinger introduces it with a page-and-a-half biography of Kummer.

The biography is rather dry. It contains the basic facts, but the choices are sometimes odd. I can't believe, for example, that many readers will be interested to know that Kummer's father's full name was Carl Gotthelf Kummer; on the other hand, they might well be interested to know that Carl Gotthelf died when his son Ernst Eduard was only 3 years old, a fact that is not mentioned. Another peculiar choice is the decision to mention that Kummer was in military service at the time that he sent his paper on the hypergeometric series to Jacobi. This seems worth mentioning only in connection with two amusing quotes. Kummer was a mere musketeer in the military, and, as he wrote to his mother, "he [Jacobi] will be amazed and at the same time pleased that a musketeer is dealing with the same matters that he is." Jacobi's reaction fulfilled his expectations. He is reported to have said, "If even musketeers now produce such outstanding mathematical works, I would like to have a look at what the noncommisioned officers are doing." Neither line is quoted here.

There are also two mathematical misstatements. It is said that Kummer used his theory of ideal prime factors "to prove the law of biquadratic reciprocity" and that he "discovered the fourth-order surface". In fact, the law of biquadratic (fourth power) reciprocity was proved (by others) before ideal prime factors were invented; Kummer used his theory to prove a law of $p$th power reciprocity for regular primes $p$ (although more than 10 years of strenuous effort came between the invention of ideal prime factors and the proof of the reciprocity law). And, of course, it was a *particular* fourth-order surface—the Kummer surface—that Kummer discovered.

The text of Kummer's article itself is given in the translation of T. F. Cope from D. E. Smith's *Source Book in Mathematics* of 1929. A few mistakes from the Smith volume are faithfully reproduced ($m\lambda - 1$ instead of $m\lambda + 1$, $u$ instead of $\mu$ in one crucial place, "number" instead of "numbers" in another, and a serious mistranslation which says "every" where it should say "some", making an important idea unintelligible). The translation does not show good understanding of the piece on the part of the translator, and contains at least one phrase—"the theory of a form of the second degree in two variables with determinant, however, a prime number $\lambda$, is

closely interwoven with these investigations"—which couldn't possibly be understood except by reference to the German original. I sympathize with Calinger here; the editor of an anthology can hardly be expected to deal with subtle shortcomings of his sources. Calinger must be blamed, though, for his failure to identify the authors of the four notes at the end of the text. (The first note is Kummer's, the other three were made by the translator.)

These shortcomings should not be allowed to obscure the main fact: This paper of Kummer, which I consider to be of major significance in the history of mathematics, is here available, albeit in slightly garbled form, to readers of English. With the aid of the German original, a serious reader could surely use it to gain a good understanding of Kummer's seminal ideas and his discussion of what motivated them. And I would like to think that the inclusion of the paper in this book will increase the likelihood that serious readers may do so.

Two minor points: First, the picture of Newton on the cover and on the first page of the book is terrible; an original drawing, it makes him look like a boxer on the day after a rough bout. The drawings of Gauss and (Emmy) Noether are no better likenesses than the one of Newton, but at least they don't make their subjects look like monsters. (As for the remaining drawing, that of Archimedes, it is of course wholly imaginary. It is innocuous enough.) Second, the title of selection #94 "From the Testamentary Letter Sent to Auguste Chevalier (May 29, 1832)" might well send readers to the original to see what has been omitted, but they will be disappointed—the selection in fact contains the entire letter.

One virtue of the book that deserves special mention is its coverage of modern mathematics. Because logic and set theory are the only important subjects in modern mathematics that are accessible to a broad spectrum of mathematicians, most history books tend to give the impression that mathematics since 1890 has been dominated by Cantor and Gödel. While this book has a fair amount of this sort of thing, it presents other topics too—Hilbert's problems, a passage from Lebesgue on integration, an excerpt of G. D. Birkhoff on the ergodic theorem, and a paper by Hasse, R. Brauer, and E. Noether on the theory of algebras (this last paper in an original translation by Kurt Bing). Not many readers will be able to follow the details of these works, but they may be able to get something from them; in any case they will learn that 20th century mathematics is not all about the axiom of choice and the continuum hypothesis.

Though not without shortcomings, this is a rich and useful book that deserves to be in every institutional mathematics library and will be a welcome addition to many private libraries (including mine).

---

*Linear Algebra through Geometry*. By Thomas Banchoff and John Wermer. Undergraduate Texts in Mathematics, edited by R. W. Gehring and P. R. Halmos. Springer-Verlag, New York, 1983. x + 257 pp.

Morris W. Hirsch
*Department of Mathematics, University of California, Berkeley, CA 94720*

I can still recall my delight—after suffering through a computational course on matrices—at learning (I think from Halmos' book *Finite-Dimensional Vector Spaces*) the geometrical and conceptual proof of the diagonalization of Hermitian operators. Students who read this book will see (among other things) the essence of that proof.

Geometry, in a broad sense, is a realm of sharply defined, thing-like mathematical objects—surfaces, spaces, sets—that we can see, if only with the mind's eye. Other parts of mathematics seem—to the geometrically oriented—to be somewhat more vague and harder to grasp: processes, relations, operations, equations, functions. Even these have become increasingly geometrified: it is much easier (for me) to understand a function as a *set* of ordered pairs than as a "rule" or collection of formulas. In fact the development of modern mathematics is to a great extent the enormous expansion of the geometric realm.

At the opening of the nineteenth century geometry meant the study of physical space, which was taken to be Euclidean. A few towering geniuses ventured into other geometries. Gauss confined his discussion of the $n$-cube to private correspondence, and his non-Euclidean geometry to his notebook. Cauchy hesitantly wrote of an $n$-tuple of numbers as an "analytic point", while Cayley wrote about $n$-dimensional geometry but explicitly denied making any assertions as to the existence of such spaces. To lesser mortals, however, space was flat and three dimensional. Möbius, for example, puzzled by the relation between an asymmetric tetrahedron and its mirror image, wrote that they were not congruent because to move one onto the other would require more than three dimensions, and "such a space is unthinkable".

The great geometrical liberation that took place during the latter half of the nineteenth century was marked by three events: the popularization of non-Euclidean geometry; Riemann's foundations of differential geometry, and especially his crucial distinction between physical and mathematical geometry; and the development of set theory. With set theory it became possible to talk about and name hitherto unspeakable things, such as the *set* of all $n$-tuples.

Along with the growth of geometry *per se* came the geometrization of other mathematical concepts. Complex function theory, for example, acquired a permanent geometric component through Riemann's introduction of the surfaces named after him. As a result, the murky process of analytic continuation became visualizable and analyzable by drawing curves on surfaces. In this century a good part of analysis became understandable—and computable—in terms of the geometry of Hilbert space.

At the same time algebra was also developing. Like other branches of mathematics, it was seen more and more in terms of sets and mappings. While Peacock and others saw abstract algebra as merely codifying operations on numbers, Hamilton insisted that algebra could reflect any mental operations whatever. As a result new systems could be invented that did not obey the same laws as numbers—for example, his non-commutative quaternions.

If geometry lets us see what we are thinking about, algebra enables us to talk precisely about what we see, and above all to calculate. Moreover it tends to organize our calculations and to conceptualize them; this in turn can lead to further geometrical construction and algebraic calculation.

A simple example, studied in the book under review, is the set of relationships among systems of linear equations, matrices, and linear transformations. Solving a system of equations is a typical example of what I mean by a "process"—we don't say what it *is*, it is not an object, we just go ahead and *do* it. Along the way we do other processes, such as "substitution" and "cancellation". There is nothing wrong with doing these things, but it is not easy to study systematically what we are doing. One way of studying this process has become part of the field of numerical analysis; but there is a need for a more structural study, aimed at answering questions such as "how many solutions are there?" and "how can we know before we try to solve whether there is in fact a solution?"

Such questions have been answered by the theory of matrices. (I like to think of matrices as geometric objects—they have diagonals, rows, columns, symmetries, etc.) The processes of solving become precisely defined operations on matrices. In particular, substitution becomes matrix multiplication.

An even better way to think about equations is in terms of mappings, more precisely, linear transformations. The product of two matrices now represents the composition of mappings. With this interpretation, associativity of matrix multiplication is simply a special case of the trivial associative law for mappings. This is much better than the direct computational proof, not so much because the latter is tedious, but rather because it affords no conceptual insight into why the law is true (other than that a proof exists).

This proof for $2 \times 2$ matrices is given by the authors in Chapter 2.3 (one of their innovations is the use of fractional chapter numbers). Even better, it is given simultaneously with the proof of associativity for linear transformations of the plane. This exemplifies the authors' approach: "We lead the student to an understanding of elementary linear algebra by emphasizing the geometrical

aspects of the subject."

Except for two brief chapters the topics are all in dimension 3 or less. They include dot and cross products, determinants, matrices, isometries, eigenvalues, conics, quadrics, linear transformations.

The book is *very* elementary. Complex numbers are not mentioned until the last page. Determinants are defined only for $2 \times 2$ and $3 \times 3$ matrices. The theorem that every subspace of $R^n$ has a well-defined dimension is stated but not proved. The concept of rank is not mentioned.

What the book covers is done well. The discussions of geometric concepts and their relations to algebraic notions are clear and well motivated. The high point is the proof of the spectral theorem for $3 \times 3$ matrices. This is applied to the classification of quadric surfaces. (The authors begin the proof, appropriately, with a geometrical appeal to the graph of a cubic to obtain a real eigenvalue.)

For my taste the book is not geometrical enough, however. For example, while diagonal, permutation and shear matrices are defined, there is no discussion or picture of the linear transformations to which they correspond. There is no discussion of similarity transformations, except orthogonal ones for diagonalizing symmetric matrices. This precludes the important insight that similar matrices correspond to the same linear transformation seen in different coordinate systems. Systems of equations are interpreted through matrices, but are not explicitly discussed in terms of linear transformations.

The greatest lack is the complete exclusion of complex numbers. Complex roots of the characteristic polynomial are not mentioned, despite their clear geometrical interpretation as rotation angles for orthogonal transformations.

Of course, these are primarily questions of taste. Since the book has evidently been carefully and skillfully planned, the authors must have deliberately excluded many topics for the sake of brevity or the desire to keep matters elementary. The subjects they include are presented with admirable clarity and simplicity. Every student of mathematics should be exposed to them as early as possible—preferably in high school. In college this book would make an excellent supplementary text for an elementary linear algebra course, or the calculus course which covers similar topics.

## LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

Reading Professor Krantz' article in the MONTHLY was enough to make any red-blooded mathematician's heart swell with pride, maybe even with indignation. Computer Science (whatever *that* is) is sexy and fun while mathematics is important and difficult. *They* get their picture in *Time* magazine while we revolutionize Volvo carburetors. There they go again.

Unfortunately, when the thrill of the prose was over I realized that the article had little relation to reality. What Professor Krantz has done is to construct two "straw fields" in the image of his prejudices, and proceed to compare them. Let's take a look at what the current state of Mathematics and Computer Science really is.

First of all, Computer Science majors do not limit their training to trivial programming in a couple of languages and fussing with trivial hardware questions. As far as I can see, Computer Science programs require students to take at least two years of Calculus, a term of Linear Algebra, often Number Theory, and courses such as Compiler Design, Comparison of Languages, Artificial Intelligence and Analysis of Algorithms. These last courses are often difficult, abstract, and

require the same skills which Professor Krantz assigns to Mathematics: abstract reasoning, modelling, and problem-solving. The difference is that the students with the greatest abilities in these areas used to go into Pure Mathematics; now, a large percentage of them go into "Computer Science". Some of the students who still major in Mathematics do so because—sad to say—the difficulty and competitiveness in Computer Science is too great. Not every C.S. major is a mindless hacker; not every Math. major is capable of proving anything beyond the simplest theorems.

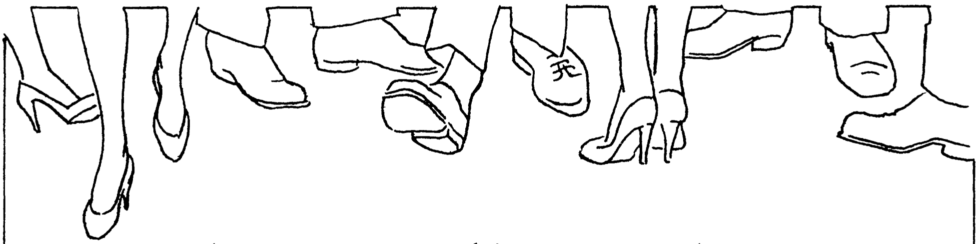
It is difficult to evaluate hearsay, non-quantified random stories such as Professor Krantz supplies us: Big Companies Want Math. Majors. What Big Companies want is students who have the creative skills that the Big Companies need. Their initial reasoning was that the attractiveness of high-tech jobs would encourage some of the best students to go into Computer Science programs. This seems to be the case. On the other hand the pressure to set up these programs and the scarcity of people competent to teach in them and willing to work in academic conditions has made the Computer Science degree an undependable indicator of ability; so has the large enrollment of students with variable ability. As usual, the business community would like the academic community to find some certification system which always works. Unfortunately this has never happened in any field at any time. Any company which thinks that a Computer Science or Mathematics degree guarantees creativity in an industrial setting is bound to be disappointed in the long run.

If we're going to talk folktales, let's recall the stories of the same Big Companies telling Mathematics Departments that, yes, many of your students are bright all right, but they can't avoid being abstract; they can't appreciate technical nitty-gritty applied work, even if the problems are interesting; they don't know enough Physics, Chemistry, Engineering, etc.; they don't *want* to work on hard technical problems. Sure, a lot of this is humbug, but it certainly should teach us to avoid stereotyping. All that Professor Krantz can legitimately say is that those Mathematics majors with the best problem-solving abilities will be of more use to the Big Companies than those Computer Science majors who are not as talented. I imagine most of us suspected this. One might try to estimate the percentage of very able students in each of the two fields, but it is difficult to measure or predict creativity; furthermore, even if we could, unless the percentages were wildly disparate it is not clear what practical conclusions could be drawn.

It is not just impolite to trivialize another person's field: it can often prove embarrassing. If Professor Krantz learned all he needed of Fortran in one afternoon, then that can only indicate that what he needed could be learned in one afternoon. My Mathematics or Computer Science students can also learn to do simple number-crunching in Fortran in one afternoon. On the other hand, my students also learn about Trees, Graphs, Hashing, Recursive Algorithms, etc.,—complicated techniques not do-able in Fortran. If Professor Krantz needed this complexity in his work, he might find one afternoon insufficient to learn it. Furthermore, in solving a problem a well-written program often can give useful results on a personal computer while a poorly-written one may need a Cray to rescue it: analyzing algorithms is a highly non-trivial subject which is central to Computer Science yet rarely dealt with by Mathematics majors; *nobody* learns it in an afternoon.

Finally, Professor Krantz seems to believe that Computer Science is somehow an unidentified field which no one has tried to analyze. All I can do in this space is refer him to Professor Donald Knuth's writings—e.g., this MONTHLY, April 1974, ten years ago, or March 1985—for enlightenment and references. Professor Knuth is a person who knows *both* fields: Mathematics and Computer Science. He sees an analogy between theorems in Mathematics and algorithms (numerical and otherwise) in Computer Science. While one field is young, the other venerable, he foresees a close collaboration for a long time. Let's not ruin it by speaking in haste and ignorance.

Mark Bridger
Department of Mathematics
Northeastern University
Boston, MA 02115



THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 92, Number 9 November 1985

## Contents

(ISSN 0002-9890)

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.




The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# RUBIK'S GROUPS

EDWARD C. TURNER
*Department of Mathematics and Statistics, SUNY at Albany, Albany, NY 12222*

KAREN F. GOLD
*Department of Mathematics, University of California at Los Angeles, Los Angeles, CA 90024*

**1. Introduction.** Few puzzles have captured the public fancy the way Rubik's cube has. The beauty of the puzzle, beyond its mechanical ingenuity and elegance and colorful appearance, lies in the contrast between the sheer impossibility of solving the cube by chance (better to bet on the proverbial snowball) and the existence of algorithms simple enough that many children master and even discover them. The value of the cube for the teacher of mathematics is that it provides a setting that is interesting in its own right and in which most of the important notions of elementary group theory are illustrated. The purpose of this article is two-fold. First we describe a uniform method—The Method—for solving *any* "Rubik's-type puzzle" (defined in the next section). As is the case with all descriptions of solution algorithms, we provide a list of basic moves, outline how to use them to unscramble a puzzle, and assure the reader that they will always suffice. Despite its generality, The Method is conceptually very simple, illustrating the fact that generalization often leads to simplification. The second purpose is to analyse the "Rubik's Group $\mathscr{R}$" whose elements are the various states of the puzzle and whose group operation describes composition of moves and to *prove* the adequacy of The Method. The analysis will involve an excursion into group theory in which permutation groups, direct sums, homomorphisms and exact sequences appear naturally and which leads to an explicit description of $\mathscr{R}$ as a semi-direct product of easily described factors. A by-product of this structure is a notation scheme that handles composition easily. We also discover that the octahedral puzzle is unique among Rubik's type puzzles in that edge flipping is impossible. In order to make the paper as accessible as possible to beginning students, we have collected in Appendix 1 a list of definitions of terms that might not be introduced in a first course in algebra. The first use of each term appears in italics.

This paper is an outgrowth of a senior research project by Gold under the direction of Turner. It began with a study of the Tetrahedron {Pyraminx} and the Cube; as more puzzles came on the market, namely the Impossiball, Alexander's Star and the Megaminx (a dodecahedron), the scope of the project widened to consideration of the general puzzle, resulting in the current analysis.

**2. The Method.** In this section we describe The Method—an algorithm that will solve any *Rubik type puzzle*. A Rubik type puzzle is a *regular solid* (see Appendix) composed of pieces corresponding to vertices, edges and faces of the solid and for which a basic move is the rotation of a face, including edge and vertex pieces. This gives 5 possibilities corresponding to the 5 regular solids, which are listed in Table 1 of the next section. The theory also applies to any puzzle whose moves are turns about a vertex—e.g., Alexander's star or the Impossiball—by the trick of dualization. Inside any regular solid is another whose vertices are the centers of the original faces (so a cube contains a smaller octahedron) and vertex turns on the larger correspond to face turns on the smaller. To solve a vertex turning puzzle, imagine the dual face turning puzzle inside and

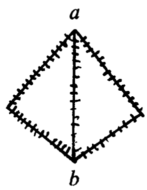
---

*Karen Gold*: I am presently a graduate student and teaching assistant in the Mathematics Department at UCLA. My primary mathematical interests are combinatorial group theory and algebraic geometry. My hobbies include Russian studies, folklore, boardgames, and the department's daily tea.

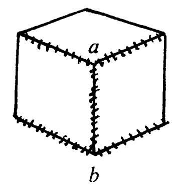
*Edward Turner*: I did my undergraduate work at the University of Rochester, graduate work at UCLA, and spent two years as an instructor at MIT before moving to SUNYA in 1971. My early mathematical interest was in differential topology, but in the last five years I have become increasingly interested in combinatorial group theory. My non-mathematical interests include running, squash, and folk dancing.

solve it. This gives another four puzzles (the dual of the tetrahedron is the tetrahedron). Each puzzle has simpler versions that don't have all three types of piece: the Impossiball has only faces, Alexander's Star only edges and the Rubik's Pocket Cube only vertices. Rubik's Revenge and obvious more general $n \times n \times n$ versions are not Rubik type puzzles in our sense. They admit another type of move not equivalent to any combination of face moves—the "slice" moves that turn a plane of pieces parallel to face.

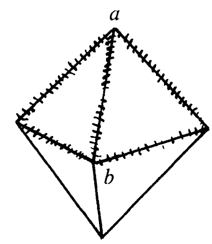
For each puzzle, consider one edge joining vertices $a$ and $b$ together with the adjacent edges on the two faces that meet at that edge, as indicated below.



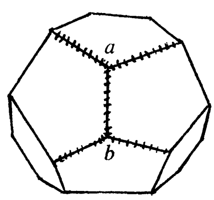

Tetrahedron



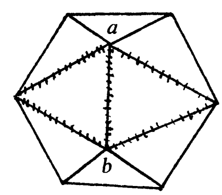

Cube




Octahedron

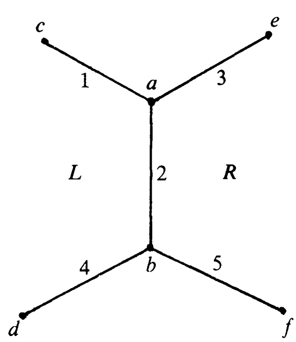



Dodecahedron




Icosahedron

We denote this schematically as shown in Fig. 1, with the understanding that $c = d$ and $e = f$ if faces are triangles and that there are other edges on top and bottom if the vertex degree is more than 3.




Small letters label vertices,
numbers edges,
and capitals face terms.

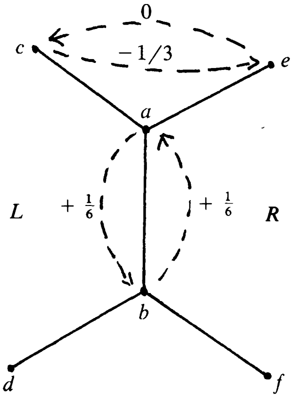
FIG. 1

We denote by $L$ and $R$ the clockwise rotations of the faces labeled $L$ and $R$ and use the convention that moves in a sequence are applied from left to right. After some experimentation with short sequences of basic moves, one discovers the usefulness of the *commutator* (see Appendix) of $L$ and $R^{-1}$, which has the effect shown in Fig. 2.

**C** is the basic building block of the solution which proceeds in four steps. The face pieces are in solved position by the nature of the puzzle and serve as references in placing vertex and edge pieces.

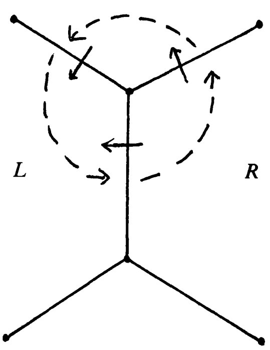
Effect of $\mathbf{C} = [L, R^{-1}] = LR^{-1}L^{-1}R$

Vertex effect




Rotation is referenced to direct parallel translation with positive fractions denoting clockwise rotations.

Edge effect




The effect of **C** on orientation is to preserve the small transverse arrows.

FIG. 2

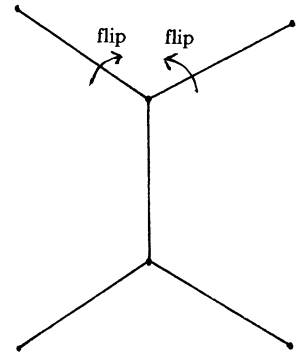
### Step 1. Place the edges

**C** is a 3-cycle on edges. By conjugation we can obtain any 3-cycle of edges; for let $T$ be any sequence that moves 1 to 4, 2 to 5 and 3 to 6 (such $T$ always exists)—then $T^{-1}\mathbf{C}T$ cycles 4, 5 and 6. Since 3-cycles generate the alternating group, we can manage any even permutation of the edges. On all puzzles except the Cube, only even permutations are possible and we are done. On the Cube, if an odd edge permutation is desired, one basic move—inducing an odd edge permutation—converts it to an even permutation. Thus **C** and its *conjugates* (see Appendix) suffice to place all the edge pieces correctly.

### Step 2. Orient the edges

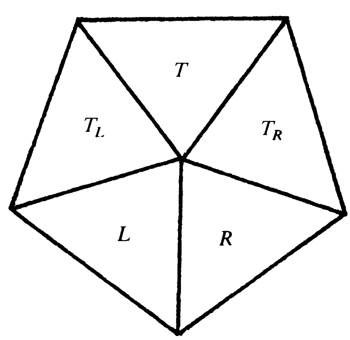
Flipping edges in pairs is particularly easy on puzzles with vertex degree 3: let $\rho$ denote clockwise rotation of the puzzle 1/3 of a turn about the diameter through vertex $a$. (Of course one never actually does the $\rho^{-1}$. This is the only time we use a rigid motion of the entire puzzle—we do so only for ease of description.) As the number of flipped edge pieces is always even, this move suffices to right them all. (See figure at the top of p. 620.)

Effect of $[\mathbf{C},\rho]=\mathbf{C}\rho\mathbf{C}^{-1}\rho^{-1}$ on edges




Edge flipping on the octahedron and icosahedron is complicated by the presence of other edges on top; we show in Lemma 2 of Section 3 that edge flipping on the octahedron is in fact impossible. A double edge flipper for the dodecahedron is shown below. (See Fig. 3.)

Dodecahedron




$[L,R^{-1}]T_L^{-1}T^{-1}[R^{-1},T_R]TT_L$

FIG. 3

After steps 1 and 2 have been completed, the edges are in solved position. The remaining steps leave them there.

Step 3. Place the vertices

We will see that once edges are placed, the vertex permutation must be even and that it suffices to have a move that cycles 3 vertices without effecting the edges. Let $B$ denote clockwise rotation of the bottom face—the one across edge $bd$ from $L$ in Fig. 1. Then $[\mathbf{C},B]$ permutes vertices by $(abd)$ and has no effect on edges.

Step 4. Orient the vertices

We will see that it suffices to rotate two vertices in opposite directions. The move $[\mathbf{C}^2,B]$ rotates $b$ clockwise and $d$ counterclockwise (as viewed from outside) and does not affect edges.

In the next two sections we will see that this algorithm will restore any scrambled puzzle to its original—often called pristine—state. It is surely inefficient in terms of time compared to others, but is conceptually quite simple and simple in practice in that very few conjugates of basic moves are necessary. In fact, the cube can be solved using conjugates only in Step 1. Furthermore, the list of basic moves is very short and easily remembered:

$$\mathbf{C} \qquad [\mathbf{C},\rho] \qquad [\mathbf{C},B] \qquad [\mathbf{C}^2,B]$$

(except for the exotic double edge flipper).

On puzzles lacking edges (like the Rubik's Pocket Cube and the Impossiball) steps 1 and 2 are unnecessary and on those lacking vertices (like Alexander's Star) steps 3 and 4 are unnecessary. There is, however, a complicating factor when faces are absent, as they are on all the above—it is much harder to decide what moves to make without reference face pieces. One learns from practice how to deal with this problem and we will ignore it.

**3. The Rubik's group $\mathscr{R}$.** With each regular solid is associated a Rubik's type puzzle and a Rubik's group, defined below. Table 1 gives the appropriate numbers for each puzzle and the last row sets our general notation.

| Regular Solid | Group | face degree | vertex degree | # of vertices | # of edges | # of faces |
|---|---|---|---|---|---|---|
| Tetrahedron | $\mathscr{T}$ | 3 | 3 | 4 | 6 | 4 |
| Cube | $\mathscr{C}$ | 4 | 3 | 8 | 12 | 6 |
| Octahedron | $\mathscr{O}$ | 3 | 4 | 6 | 12 | 8 |
| Dodecahedron | $\mathscr{D}$ | 5 | 3 | 20 | 30 | 12 |
| Icosahedron | $\mathscr{I}$ | 3 | 5 | 12 | 30 | 20 |
| Generic | $\mathscr{R}$ | $p$ | $q$ | $V$ | $E$ | $F$ |

TABLE 1

We imagine the central mechanism of the puzzle as fixed, so the face pieces don't move, and refer to the $F$ rotations of a face through $2\pi/p$ radians clockwise as "basic moves". An element of the group $\mathscr{R}$ is a sequence of basic moves, where it is understood that two sequences represent the same move if their effect on the puzzle is the same. Otherwise stated, $\mathscr{R}$ is the quotient of the *free group* (see Appendix) generated by the basic moves by the normal subgroup of expressions that leave all vertex and edge pieces in their original positions and orientations.

We analyse $\mathscr{R}$ by separating the position effect from the orientation effect and the vertex effect from the edge effect. Ignoring orientations, each element of $\mathscr{R}$ permutes the edges and vertices, defining a map pos (for position)

$$\mathscr{R} \xrightarrow{\text{pos}} S_E \times S_V,$$

where $S_n$ is the permutation group on $n$ letters. We denote the kernel of pos by $\mathscr{F}$, the position fixed subgroup of moves that do not move pieces from their original positions and the image of pos by $\Sigma$. Thus by definition we have an *exact sequence* (see Appendix)

$$1 \to \mathscr{F} \xrightarrow{i} \mathscr{R} \xrightarrow{\text{pos}} \sum \to 1,$$

where $i$ denotes inclusion.

The remainder of this section is devoted to identifying $\mathscr{F}$ and $\Sigma$—the analysis of $\mathscr{R}$ is completed in the next section with a description of how $\mathscr{F}$ and $\Sigma$ interact.

LEMMA 1.
(i) *If $\mathscr{R} \neq \mathscr{C}$, then $\Sigma = A_E \times A_V$.*
(ii) *If $\mathscr{R} = \mathscr{C}$, then $\Sigma = \{(\sigma, \mu) | \sigma$ and $\mu$ are both even or both odd$\}$.*

LEMMA 2. *There are exact sequences*:
(i) *If $\mathscr{R} \neq \mathscr{O}$,*

$$0 \to \mathscr{F} \xrightarrow{f \times r} \bigoplus_1^E \mathbb{Z}_2 \oplus \bigoplus_1^V \mathbb{Z}_q \xrightarrow{\text{Sum}} \mathbb{Z}_2 \times \mathbb{Z}_q \to 0,$$

*where*

$$\mathrm{Sum}((f_1,\ldots,f_E),(r_1,\ldots,r_v)) = \left(\sum_1^E f_i, \sum_1^V r_i\right).$$

(ii) *If* $\mathscr{R} = \mathcal{O}$,

$$0 \to \mathscr{F} \xrightarrow{r} \bigoplus_1^V \mathbb{Z}_2 \xrightarrow{\mathrm{Sum}} \mathbb{Z}_2 \to 0.$$

This is clearly a *split short exact sequence* (see Appendix), so this implies that $\mathscr{F}$ is isomorphic to

$$\bigoplus_1^{E-1} \mathbb{Z}_2 \oplus \bigoplus_1^{V-1} \mathbb{Z}_q.$$

The splitting is not *canonical* (see Appendix), however, and it is more natural to think of $\mathscr{F}$ as a subgroup of $\oplus_1^E \mathbb{Z}_2 \oplus \oplus_1^V \mathbb{Z}_q$.

*Proof of Lemma* 1. In case (i), the face degree $p$ is odd, so that each basic move induces an even permutation on both edges and faces: thus $\Sigma \subset A_E \times A_V$. $A_E$ is generated by 3-cycles and Step 1 of The Method shows how to achieve any desired 3-cycle of edges. Given $(\sigma,\mu) \in A_E \times A_V$, let $M$ be a sequence of basic moves so that $\mathrm{pos}(M) = (\sigma,\mu')$. Now $\mu'$ is even so $(\mu')^{-1}\mu \in A_V$ and Step 3 shows how to find a sequence of moves $N$ such that $\mathrm{pos}(N) = (\mathrm{id},(\mu')^{-1}\mu)$. Then $\mathrm{pos}(MN) = (\sigma,\mu)$, so pos is onto $A_E \times A_V$.

In case (ii), the face degree is 4, so each basic move induces an odd permutation on both edges and vertices—the total parity is even, so

$$\Sigma \subset A_E \times A_V \cup (S_E \setminus A_E) \times (S_V \setminus A_V).$$

The argument of case (i) shows that $\Sigma \supset A_E \times A_V$. If $\sigma$ and $\mu$ are both odd, then let $B$ be any basic move and consider pos $(B)\cdot(\sigma,\mu) = (\sigma',\mu')$. $\sigma'$ and $\mu'$ are both even, so $(\sigma',\mu') \in \Sigma$; thus $(\sigma,\mu) \in \Sigma$ and we are done.

*Proof of Lemma* 2. The key to Lemma 2 is the question, "How do you measure the flipping of an edge piece or the rotation of a vertex piece when it has changed position?" It may seem paradoxical to ask this question when we are studying the position fixed subgroup—we do so because we need to think of elements of $\mathscr{F}$ as products of basic moves, which do move pieces. To answer the question, consider the graph whose vertices are the vertices of the puzzle and whose edges join two vertices that can be obtained one from the other by the application of a single basic move. (In fact, these will correspond to the edges of the puzzle, but they should be thought of in this way.) Now choose a *tree* (see Appendix) containing all the vertices, and for each edge of the tree, choose a basic move that moves a vertex piece along that edge.

Shown below, in heavy lines, is a particular choice of a vertex tree for $\mathscr{C}$. (See Fig. 4.)

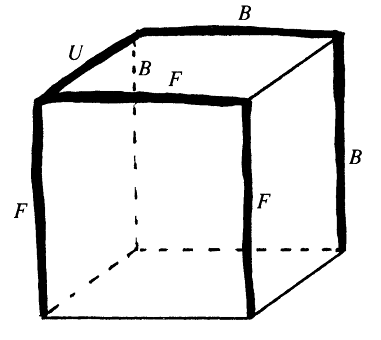



FIG. 4

The letters are the standard *Singmaster notation* (see Appendix). A tree has the property that there is only one way to get from one vertex to another in the tree. The answer to the key question, for vertex pieces, is then "compare the effect of the move in question to the effect of the unique sequence of basic moves that moves that vertex along the tree using the labeling moves on each edge." This gives a standard of comparison for every possible position change. Thus, for example, to move the lower-back-right vertex of $\mathscr{C}$ to the lower-front-right position, the standard sequence is $B^2U^{-1}F^2$. The move $R$, rotation clockwise of the right face, has a different effect—namely rotation 1/3 of a turn clockwise from the standard.

On the general puzzle, for each move $M$ and each vertex piece $v$, we have an integer $r_v(M)$ (mod $q$) that measures how many $(2\pi/q)$ radian clockwise rotations are necessary to move the standard reference position for that vertex move to the one given by the move in question. In a similar manner, consider the edge graph whose vertices are the edge pieces of the puzzle and whose edges join pieces that can be gotten one from the other by a single basic move. The edge tree is any tree in the edge graph containing all its vertices—it is not necessary to label the edges of this tree since the moves are uniquely determined. Shown below is a particular choice of edge tree for $\mathscr{C}$, with labels included for convenience. (See Fig. 5.)

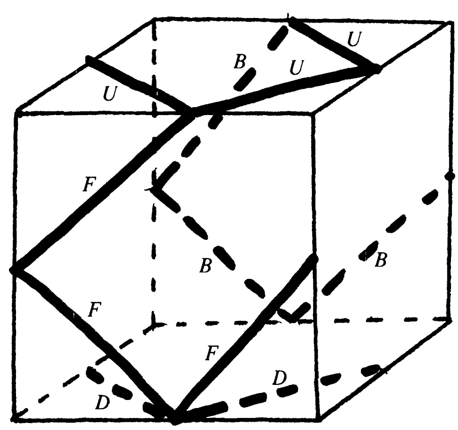



FIG. 5

Then to each move $M$ and edge piece $e$ we associate an integer $f_e(M)$ (mod 2) that measures whether or not the edge piece $e$ is flipped relative to the standard by the given move.

This defines a map

$$\mathscr{R} \xrightarrow{f\times r} \bigoplus_1^E \mathbb{Z}_2 \oplus \bigoplus_1^V \mathbb{Z}_q,$$

$$f(M) = \left(f_{e_1}(M),\dots,f_{e_E}(M)\right), \quad r(M) = \left(r_{v_1}(M),\dots,r_{v_V}(M)\right).$$

The proof of Lemma 2 for $\mathscr{R} \neq \mathscr{O}$ will be complete when we have verified the following claims.

*Claim* 1. $(f \times r)|\mathscr{F}$ is a homomorphism.

*Claim* 2. Image$(f \times r) \subset$ ker(Sum).

*Claim* 3. Image$(f \times r) \supset$ ker(Sum) if $\mathscr{R} \neq \mathscr{O}$.

Note that $f \times r$ is not a homomorphism on $\mathscr{R}$ because of the movement of edges and faces: in fact, it is easy to check that

$$(**) \qquad \begin{aligned} r_v(M_1M_2) &= r_v(M_1) + r_{\mu_1(v)}(M_2),\\ f_e(M_1M_2) &= f_e(M_1) + f_{\sigma_1(e)}(M_2), \end{aligned}$$

where $\operatorname{pos}(M_1) = (\sigma_1, \mu_1)$.

Claim 1 is clear from equations $(**)$. It is also clear that $\operatorname{Sum}(f \times r) = (\Sigma_e f_e, \Sigma_v r_v)$ is also a homomorphism, so that to verify Claim 2, it is only necessary to prove that $\operatorname{Sum}(f \times r)(B) = 0$ for any basic move $B$. Suppose $B$ is the basic move of a face depicted below (we take $p = 5$ for ease of exposition—the argument is general; see Fig. 6).

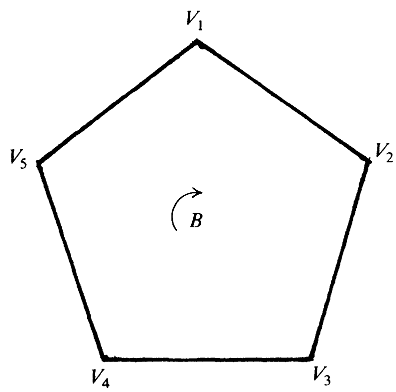



FIG. 6

The indicated vertices are the only ones moved, so

$$\operatorname{Sum}(r(B)) = r_{v_1}(B) + r_{v_2}(B) + r_{v_3}(B) + r_{v_4}(B) + r_{v_5}(B).$$

But formula $(**)$ says that the right-hand side is $r_{v_1}(B^5)$, since $B^i(v_1) = v_{1+i}$. Since $B^5 = id$, $\Sigma_{i=1}^5 r_{v_i}(B) \equiv 0 \pmod 5$.

An analogous argument works on the edges.

(REMARK. It is at this point of the argument that $(**)$ is essential—an arbitrary method of measuring twisting and flipping would be inadequate.)

The proof of claim 3 for $\mathscr{R} \neq \mathcal{O}$ is an exercise in the use of The Method that is left to the reader: it involves explicit construction of moves whose images under $f \times r$ generate ker(Sum).

In the case of the octahedron, the above analysis holds except that:

(a) No edge flipping is possible, so that $\mathbb{Z}_2$ factors measuring flips are absent;

(b) the vertex rotations must come in units of $2(2\pi/4)$, not $2\pi/4$ as expected.

Both of these can be seen by dividing the faces into two classes—say rough-textured and smooth-textured—so that each edge separates a rough face from a smooth one. The possibility of doing this is easily seen to be equivalent to even vertex degrees. Then each basic move and so all of $\mathcal{O}$ preserves the texturing. It follows that edge pieces can't have their rough and smooth facelets interchanged nor can vertex pieces be rotated by $2\pi/4$ or $3(2\pi/4)$.

## 4. The semi-direct product structure on $\mathscr{R}$.

DEFINITION. Suppose $A$ and $B$ are groups and $\varphi$: $B \to \operatorname{Aut}(A)$ is a homomorphism. Then the *semi-direct product* of $A$ and $B$ on $\varphi$ is $A \times B$ as a set with the product

$$(a, b)(a', b') = (a\varphi(b)(a'), bb').$$

It is routine to check that this gives a group structure. It agrees with the direct product if and only if $\varphi$ is the trivial homomorphism. The following standard theorem describes three equivalent ways to express this notion: see for example, W. R. Scott, *Group Theory*, Prentice-Hall, NJ, 1964, p. 213.

PROPOSITION. *The following are equivalent*:

(i) *There is a split short exact sequence*

$$1 \to A \xrightarrow{\alpha} G \underset{\psi}{\overset{\beta}{\rightleftarrows}} B \to 1.$$

(ii) *$G \simeq A \times_\varphi B$ for some $\varphi$.*
(iii) *$G$ has two subgroups $A$ and $B$ so that*

$$A \lhd G, A \cap B = \{\text{id}\} \quad \textit{and} \quad G = AB.$$

REMARK. Several $\psi$'s may correspond to the same $\varphi$: if $\psi_2(b) = \psi_1(b)c$, where $c \in$ Centralizer of $A$, then $\psi_1$ and $\psi_2$ will determine the same $\varphi$. In particular, this will happen if $A$ is abelian and $c \in A$. This illustrates how $\varphi$ may be canonical in a setting in which $\psi$ is not.

We now have the tools to complete the analysis of $\mathscr{R}$. Assume edge and vertex trees, with labels, have been chosen and let $\psi$: $\Sigma \to \mathscr{R}$ by

$$\psi(\sigma, \mu) = M,$$

where $\text{pos}(M) = (\sigma, \mu)$ *and* $f_v(M) = r_p(M) = 0$ for all $v$ and $e$. $(**)$ implies that $\psi$ is a homomorphism. Corresponding to $\psi$, we have $\varphi$: $\Sigma \to \text{Aut}(\mathscr{F})$ by

$$\varphi(\sigma, \mu)((f_1, \dots, f_E), (r_1, \dots, r_V)) = (\sigma(f_1, \dots, f_E), \mu(r_1, \dots, r_V)),$$

where $\sigma(f_1, \dots, f_E)$ and $\mu(r_1, \dots, r_V)$ are the sequences obtained by permuting the entries according to $\sigma$ and $\mu$.

THEOREM. (i) *The following is a split exact sequence*:

$$1 \to \mathscr{F} \to \mathscr{R} \underset{\psi}{\overset{\text{pos}}{\rightleftarrows}} \Sigma \to 1.$$

(ii) $\mathscr{R} \simeq \mathscr{F} \times_\varphi \Sigma$.
(iii) *$\mathscr{R}$ has subgroups $\mathscr{F}$ and $\psi(\Sigma)$ such that $\mathscr{F} \lhd \mathscr{R}$, $\mathscr{F} \cap \psi(\Sigma) = \{\text{id}\}$ and $\mathscr{F}\Sigma = \mathscr{R}$.*

*Proof*. By the proposition, it suffices to prove any one of the three and to check that $\psi$ and $\varphi$ are correctly related. That $\psi$ splits pos is immediate, verifying (i). The proposition says that $\varphi(\sigma, \mu)$ $(f)$ should "permute the edges and faces by $\sigma$ and $\mu$ without flips or twists—then flip and twist according to $f$—then restore edges and faces to their original positions without flips or twists." This is clearly just what $\varphi(\sigma, \mu)$ $(f)$ does. We point out that $\psi$ was not canonical, depending on the choice of tree, but $\varphi$ *is* canonical.

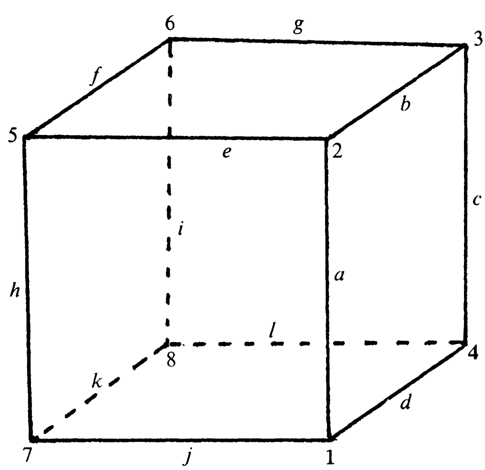



FIG. 7

The theorem gives an explicit description of $\mathscr{R}$ which provides a convenient notation that we illustrate for Rubik's Cube. We label the vertex and edge pieces of **C** with numbers and letters

respectively as below, relative to which the permutations induced by the basic moves in Singmaster's notation can be easily determined. (See Fig. 7.)

| Move | Value of pos |
|---|---|
| $R$ | $((a\ b\ c\ d),\ (1\ 2\ 3\ 4))$ |
| $L$ | $((i\ f\ h\ k),\ (5\ 7\ 8\ 6))$ |
| $U$ | $((b\ e\ f\ g),\ (3\ 2\ 5\ 6))$ |
| $D$ | $((d\ l\ k\ j),\ (1\ 4\ 8\ 7))$ |
| $F$ | $((a\ j\ h\ e),\ (2\ 1\ 7\ 5))$ |
| $B$ | $((c\ g\ i\ l),\ (4\ 3\ 6\ 8))$ |

Now using the vertex and edge trees for **C** given in §3, we can determine the flip or rotation effect on each piece moved—the effects are filled in below in the gap representing the piece move with the absence of a label indicating zero.

| Move | Extended notation |
|---|---|
| $R$ | $((a^1\ b^1\ c^1\ d^1),\ (1^1\ 2^1\ 3\ 4^1))$ |
| $L$ | $((i^1\ f^1\ h^1\ k^1),\ (5^1\ 7^1\ 8^1\ 6))$ |
| $U$ | $((b\ e\ f\ g),\ (3\ 2^1\ 5^1\ 6^1))$ |
| $D$ | $((d\ l\ k\ j),\ (1^1\ 4^1\ 8\ 7^1))$ |
| $F$ | $((a\ j\ h\ e),\ (2\ 1\ 7\ 5))$ |
| $B$ | $((c\ g\ i\ l),\ (4\ 3\ 6\ 8))$ |

Clearly the extended notation determines the move completely. Furthermore, composition of moves is described by multiplying permutations in the usual way (left one applied first) and filling each gap with the sum of the numbers in the gaps that give rise to it. For example

$$\begin{aligned} RU &= \big((a^1\, b^1\, c^1\, d^1),(1^1\, 2^1\, 3\, 4^1)\big)\cdot\big((b\, e f g),(3\, 2^1\, 5^1\, 6^1)\big) \\ &= \big((a^1\, e f g\, b^1\, c^1\, d^1),(1^2\, 5^1\, 6^1\, 3\, 4^1)(2^1)\big) \end{aligned}$$

where, e.g.,

$$1 \xrightarrow[R]{1} 2 \xrightarrow[U]{1} 5 \Rightarrow 1 \xrightarrow[(RU)]{2} 5.$$

This notation allows us to easily draw a non-obvious conclusion. Note first that it is a relatively easy exercise to check that any move can be effected without using one of the six moves—The Method, for example, can be used to show how to obtain the effect of $F$ from the other 5.

CLAIM. *It is not generally possible to solve the cube using only* 4 *of the* 6 *moves.*

*Easy proof.* If the two stationary faces are adjacent, the edge piece between them can't be moved, and the claim is obvious. If not, they are opposite and we may assume, without loss of generality, that they are $R$ and $L$. But a glance at the chart shows that this makes edge flipping relative to the chosen tree impossible. Q.E.D.

**5. Some group theory related to the method.** In this section we describe some interesting subgroups of $\mathscr{R}$ generated by basic commutators which provide *almost* all the moves needed to solve a puzzle. We encounter two groups $K$ and $I$—discussed below—that are interesting in their own right. Verification of statements not explicitly discussed is tedious but routine.

DEFINITION. For group $\mathscr{R}$ and a degree $q$ vertex $v$ of the corresponding solid, $\mathscr{R}_v$ is the

subgroup generated by basic commutators of the form $\mathbf{C} = [L, R^{-1}]$ for the $q$ pairs of adjacent faces containing $v$. For an edge $e$, $\mathscr{R}_e$ is the subgroup generated by all commutators of moves of the two faces separated by $e$—namely $[L, R]$, $[L^{-1}, R]$, $[L, R^{-1}]$ and $[L^{-1}, R^{-1}]$.

*The groups* $\mathscr{R}_v$

$\mathscr{R}_v$ moves $q$ edges and $q+1$ vertices. In all cases, any edge flip, corner rotation or even edge permutation of these pieces possible in $\mathscr{R}$ is also possible in $\mathscr{R}_v$—the interest lies in the realizable vertex permutations.

*Case* (i). $\mathscr{R} = \mathscr{T}, \mathscr{C}, \mathscr{D}$ $(q = 3)$; Image(pos) $= A_3 \times K$.

*Case* (ii). $\mathscr{R} = \mathscr{O}$ $(q = 4)$; Image(pos) $= A_4 \times A_5$.

*Case* (iii). $\mathscr{R} = \mathscr{I}$ $(q = 5)$; Image(pos) $= A_5 \times I$.

*The groups* $\mathscr{R}_e$

If faces have degree $p = 3$ (respectively $p = 4, 5$) then $\mathscr{R}_e$ moves 5 edges and 4 vertices (respectively 5 edges and 6 vertices). $\mathscr{R}_e$ has no flip or rotation effect (pick the right tree) so $\mathscr{R}_e$ is isomorphic to Image(pos).

*Case* (i). $\mathscr{R} = \mathscr{T}, \mathscr{O}, \mathscr{I}$ $(p = 3)$ Image(pos) $= A_5 \times K$.

*Case* (ii). $\mathscr{R} = \mathscr{C}, \mathscr{D}$ $(p = 4, 5)$ Image(pos) $= A_5 \times I$.

*The Klein four group K*

$K$ is the four element subgroup of $A_4$ consisting of double transpositions. That double transpositions do not generate $A_4$ (as they do $A_n$ for $n \geqslant 5$) indicates why it is necessary to bring in the move $B$ in Step 3 of The Method. Abstractly, $K$ is just $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

*The icosahedral group I*

$I$ is the subgroup of $A_6$ generated by the two 5-cycles $X = (12345)$ and $Y = (16235)$ obtained as follows. For $\mathscr{R}_v$, case (iii), label vertices as shown in Fig. 8 and let $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_5$ be the basic commutators corresponding to edges 1–6, 2–6, …, 5–6; then

$$X = (12345) = \mathbf{C}_4^{-1}\mathbf{C}_1\mathbf{C}_2\mathbf{C}_4$$

$$Y = X^{-1}(\mathbf{C}_1\mathbf{C}_5)^2 X.$$

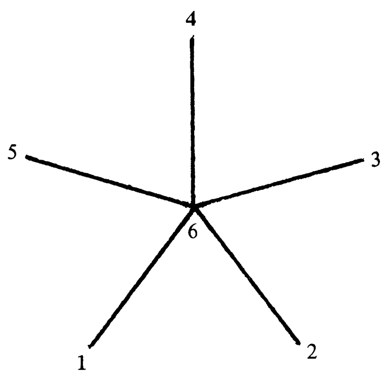
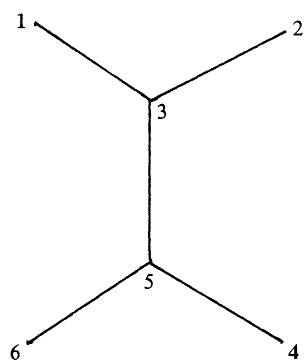



FIG. 8




$\mathbf{A} = [L, R]$

$\mathbf{B} = [L^{-1}, R]$

$\mathbf{C} = [L, R^{-1}]$

$\mathbf{D} = [L^{-1}, R^{-1}]$

FIG. 9

For $\mathscr{R}_e$, case (ii), label vertices as shown in Fig. 9; then $X = \mathbf{CA}$, $Y = (\mathbf{CD})^2$.

$I$ is a very famous group which occurs naturally in a number of different settings and is known by names that reflect its various origins. It has order 60 and is the smallest example of a non-abelian simple group (see Appendix). We end with four descriptions of $I$ other than as Image(pos).

(1) $I \simeq \langle X, Y | X^5 = Y^5 = (XY)^3 = (Y^{-1}X)^2 = 1 \rangle.$

This notation, from combinational group theory, means that $I$ is a group characterized by the fact that it has two generators $X$ and $Y$ satisfying the listed relations and such that any other relations are derivable from these together with the group axioms. Descriptions of this type, called presentations, are completely general and very efficient but rarely display the special characteristics that make a group interesting.

(2) $I \simeq A_5$, *the alternating group on* 5 *letters*.

This is the most familiar form of $I$. Here

$$X = (12534),$$
$$Y = (13452).$$

In this form, it is clear that $I$ has elements of orders 1, 2, 3 and 5 only and it's easy to count how many of each there are. Furthermore, all elements of order 2 are conjugate, all of order 3 conjugate and there are two conjugacy classes of elements of order 5.

(3) $I \simeq \mathrm{PSL}(2,5)$.

$$\mathrm{SL}(2,5) = \left\{ \begin{pmatrix} ab \\ cd \end{pmatrix} \middle| a, b, c, d \in \mathbb{Z}_5, ad - bc = 1 \right\}.$$

$$\mathrm{PSL}(2,5) = \frac{\mathrm{SL}(2,5)}{\{\pm I\}}.$$

Here

$$X = \left[ \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix} \right] \quad \text{and} \quad Y = \left[ \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \right].$$

(4) $I \simeq$ *the symmetry group of the icosahedron*.

In this incarnation, $I$ is the group of rigid motions of the icosahedron—i.e., length and angle preserving linear maps (orthogonal maps) of $\mathbb{R}^3$ that carry an icosahedron centered at the origin back onto itself. Every orthogonal map of $\mathbb{R}^3$ is a rotation about some straight line through the origin and those in the icosahedral group are of three types depending on whether the line passes through the center of an edge, the center of a face or a vertex. If $L$ and $R$ denote adjacent faces as well as clockwise rotation about the line through the center of the faces, then

$$X = R,$$
$$Y = L^2.$$

It is a delightful coincidence that the icosahedral group appears in the analysis of the icosahedral puzzle.

**Appendix 1—Definitions.**

*Regular Solid.* A regular or Platonic solid is a convex solid bounded by planar faces each of which is a regular polygon with the same number of edges and such that the same number of faces meet at each vertex. It was known to the ancient Greeks that Table 1 lists all possibilities—a proof appears in Book 13 of Euclid's Elements.

*Commutator.* The commutator of $x$ and $y$ in a group $G$ is $[x, y] = xyx^{-1}y^{-1}$. (Some authors

use $[x, y] = x^{-1}y^{-1}xy$.) It measures the degree to which $x$ and $y$ fail to commute with one another.

*Conjugate*. In a group $G$, $a$ is conjugate to $b$ if there is a $c$ so that $a = cbc^{-1}$. Conjugate permutations always have the same cycle structure.

*Free group*. A free group is a group free of any relations not demanded by the group axioms. A group free on symbols $X_1, \ldots, X_n$ is the set of all strings—called words—of $X_i$'s with exponents $\pm 1$ in which occurrences of $X_i X_i^{-1}$ and $X_i^{-1} X_i$ have been deleted, together with 1 interpreted as the empty word. The multiplication is just juxtaposition followed by the above deletions. The discipline of combinatorial group theory views all groups as quotients of free groups as in the first description of $I$ in Section 5.

*Exact sequence*, *short exact sequence*. An exact sequence is a sequence of group homomorphisms such that the image of each is the kernel of the next. A short exact sequence (the only kind considered in this paper) is 5 terms long, beginning and ending with trivial groups.

$$1 \to A \xrightarrow{\alpha} G \xrightarrow{\beta} B \to 1.$$

If all groups in the sequence are abelian, denote the trivial group by 0: otherwise 1.

*Split short exact sequence*. A splitting of the exact sequence above is a map $\psi$: $B \to G$ such that $\beta \circ \psi(b) = b$. If $G$ is abelian, the existence of a splitting says that $G$ is the direct sum of $A$ and $B$: this follows from the Proposition of §4.

*Tree*. A graph is a tree if it is connected and contains no cycles. This is the same (for finite graphs) as saying that the number of vertices is one more than the number of edges. An important property of a tree is that there is exactly one way to get from one vertex to another in the tree without retracing.

*Singmaster notation*. Singmaster's notation—which has become standard—denotes the basic moves on the Rubik's cube by letters indicating the face being rotated clockwise (see Fig. 10):

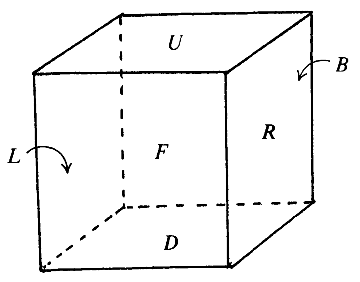



FIG. 10

| | | |
|---|---|---|
| $L$ = left | $U$ = up | $F$ = front |
| $R$ = right | $D$ = down | $B$ = back. |

(*Note*: $L$ and $R$ here should not be confused with $L$ and $R$ in The Method).

*Canonical*. A map or construction is canonical if it does not depend on arbitrary choices. In §3, $\varphi$ is canonical but $\psi$ is not, since $\psi$ depends on the trees chosen.

*Simple group*. A group is simple if it has no non-trivial normal subgroups, or equivalently, no non-trivial quotients.

His main work is invariant under translation. (See p. 661.)

The screening time per question worked becomes $\dfrac{\gamma}{\int_0^\tau g(x)\,dx}$. Thus

$$\phi(\tau) = \frac{\gamma + \int_0^\tau xg(x)\,dx}{\int_0^\tau g(x)\,dx},$$

$$\phi'(\tau) = \frac{\tau g(\tau)\int_0^\tau g(x)\,dx - g(\tau)\left(\gamma + \int_0^\tau xg(x)\,dx\right)}{\left[\int_0^\tau g(x)\,dx\right]^2}$$

$$= (\tau g(\tau) - \phi(\tau)g(\tau))\Big/\int_0^\tau g(x)\,dx$$

$$\phi'(\tau) = (-g(\tau)(\phi(\tau) - \tau))\Big/\int_0^\tau g(x)\,dx.$$

Since $\gamma > 0$, $\lim_{\tau\to 0}\phi(\tau) = +\infty$ and $\lim_{\tau\to\infty}\phi(\tau) = \alpha + \gamma$. Also the graph of $\phi$ cannot cross the diagonal from below. It follows that $\phi$ has a unique fixed point and it occurs at the min of $\phi$.

We now restate our answer to the question posed at the beginning. If, after screening, the estimate of the time to work the problem is longer than the average time spent per problem worked, then move on to the next problem. This answer may not be terribly helpful to the test taker and so we present in a table the results of the numerical calculations for the two models above for a 60 question three-hour exam with $E(T) = 6$.

| | $E(S)$ | $E(W)$ | Best $\tau$ | Answers | Screens |
|---|---|---|---|---|---|
| Chan model | 1 | 5 | 3.534 | 50.9 | 100.57 |
| | 1.77 | 4.23 | 4.561 | 39.47 | 60 |
| | 2 | 4 | 4.793 | 37.5 | 53.83 |
| | 3 | 3 | 5.524 | 32.6 | 38.71 |
| | 4 | 2 | 5.895 | 30.5 | 32.24 |
| Alternative model | 1 | 5 | 4.367 | 41.2 | 79.13 |
| | 1.4 | 4.6 | 4.855 | 37.1 | 60 |
| | 2 | 4 | 5.358 | 33.6 | 44.9 |
| | 3 | 3 | 5.818 | 30.9 | 34.45 |
| | 4 | 2 | 5.980 | 30.1 | 30.64 |

Thus, the first model suggests that, on this test, a problem should be screened for 1.77 minutes and then if it is estimated that the time needed to work the problem is more than 4.561 minutes, go on to the next. The second suggests screening for 1.4 minutes with a moving-on time of 4.855 minutes.

**Reference**

1. Beda Chan, Moving On, The Actuary, vol. 18, no. 1, Jan. 1984, 4–5.

---

**ANSWER TO PHOTO ON PAGE 630**

Alfred Haar (1885–1933).

# ON THE SECOND DERIVATIVE TEST FOR CONSTRAINED LOCAL EXTREMA

D. SPRING
*Department of Mathematics, York University, Downsview, Ontario, Canada M3J 1P3*

**1. Introduction.** A well-known mathematical problem which arises in numerous mathematical and applied contexts is the determination of the local extrema (i.e., the local maxima or local minima) of a function which is subject to subsidiary constraints. Let $f, g_\alpha : U \to \mathbb{R}$ be differentiable functions, $1 \leqslant \alpha \leqslant m$, where $U$ is open in $\mathbb{R}^n$ and $m < n$. The *constraints* are defined to be the equations $g_\alpha = 0, 1 \leqslant \alpha \leqslant m$. Let $A \subset U$ be the solution set of the constraints. The problem is to determine the local extrema of the function $\tilde{f}: A \to \mathbb{R}$, where $\tilde{f}$ is the restriction of the function $f$ to the solution set $A$. To illustrate, let $(x_1, \ldots, x_n)$ be coordinates in $\mathbb{R}^n$ and consider the *simple constraints* $x_\alpha = 0, 1 \leqslant \alpha \leqslant m$. The solution set is $A = U \cap V$, where $V$ is the linear subspace of $\mathbb{R}^n$ obtained by setting the first $m$ coordinates equal to zero. Setting $\tilde{f}(x_{m+1}, \ldots, x_n) = f(0, \ldots, 0, x_{m+1}, \ldots, x_n)$, then subject to these constraints we find that $f$ has a local extremum at $a = (0, \ldots, 0, a_{m+1}, \ldots, a_n) \in A$ if and only if the function $\tilde{f}$ has a local extremum (in the usual sense) at $\tilde{a} = (a_{m+1}, \ldots, a_n)$. In particular, $\tilde{a}$ is a *critical point* of the function $\tilde{f}$, i.e., all the first order partial derivatives of $\tilde{f}$ vanish at $\tilde{a}$. For general constraints the function $\tilde{f}$ is known only implicitly; hence the local extrema of $\tilde{f}$ cannot be calculated directly.

The classical approach to solving the problem of constrained extrema involves the well-known *method of Lagrange multipliers* [4], [6]. Let $L: \mathbb{R}^m \times U \to \mathbb{R}$ be the Lagrangian function of $(m + n)$ variables:

$$L(\lambda_1, \ldots, \lambda_m, x_1, \ldots, x_n) = f(x_1, \ldots, x_n) + \sum_{\alpha=1}^{m} \lambda_\alpha g_\alpha(x_1, \ldots, x_n).$$

Fix $a \in A$ and suppose for technical reasons, the map $g = (g_1, \ldots, g_m): U \to \mathbb{R}^m$ is of class $C^1$ and has maximal rank ($= m$) at $a$. This hypothesis ensures that, near $a$, the solution set $A$ is an $(n - m)$-dimensional submanifold of $\mathbb{R}^n$. A principal result is the following [4], [11]: Subject to the constraints, if $f$ has a local extremum at $a \in A$, then for some vector $c = (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$, the vector $v = (c, a) \in \mathbb{R}^m \times U$ is a critical point of the Lagrangian $L$. That is, all the first order partial derivatives of $L$ vanish at $v$:

$$(1.1) \qquad \frac{\partial f}{\partial x_j}(a) + \sum_{\alpha=1}^{m} c_\alpha \frac{\partial g_\alpha}{\partial x_j}(a) = 0, \quad 1 \leqslant j \leqslant n; \quad g_\alpha(a) = 0, \quad 1 \leqslant \alpha \leqslant m.$$

Generically then, the solution of the problem of constrained extrema is to be found among the critical points of the associated Lagrangian function.

However, the critical point condition (1.1) is only a necessary condition—not a sufficient condition, in general, for a constrained local extremum. The determination of appropriate sufficient conditions is the central problem addressed in this paper.

It is convenient to record, for reference, the standard approach to sufficient conditions. In the unconstrained case, $m = 0$, $L = f$ and $a \in U$ is a critical point of $f$: $\dfrac{\partial f}{\partial x_i}(a) = 0, 1 \leqslant i \leqslant n$. Assuming $f$ is of differentiability class $C^3$, one expands $f$ in a Taylor Series about $a$ ($h = (h_1, \ldots, h_n) \in \mathbb{R}^n$):

*David Spring*: The problem of an optimal second derivative test, addressed in this paper, came to my attention during a recent teaching assignment of Calculus at York University. I received my Ph.D. in Differential Topology (1967) at the University of California, Berkeley, under the direction of M. Hirsch. I have taught also at the University of Toronto and, during a seven-year period in France, at the University of Paris, Orsay, and the University of Provence, Marseilles. Current interests include topological existence theorems for solutions to non-linear P.D.E.'s. I enjoy folk and dance music, reggae music in particular.

$$f(a+h)=f(a)+\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(a)h_ih_j+R_2(h), \tag{1.2}$$

where $\lim_{h\to 0}\frac{R_2 h}{\|h\|^2}=0$. The character of the critical point $a$ is analysed by examining the *quadratic form*

$$q(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 f}{\partial x_i\partial x_j}(a)h_ih_j$$

that appears in (1.2). Note that the matrix associated to $q$ is the *Hessian matrix*

$$Hf(a)=\left[\frac{\partial^2 f}{\partial x_i\partial x_j}(a)\right]_{1\leqslant i,j\leqslant n}.$$

The remainder term is "negligible", so the following result obtains (Edwards [**6**, page 138]):

PROPOSITION 1.1. *If $q$ is positive* (*negative*) *definite, then $f(a)$ is a strict local minimum* (*maximum*) *value; if $q$ attains both positive and negative values, then $f$ does not have a local extremum at $a\in U$.*

REMARK 1.1. In the remaining cases, i.e., $q$ is degenerate ($\det Hf(a)=0$) and is positive or negative semi-definite, the character of the critical point $a$ is necessarily indeterminate at the level of second order partial derivatives (for a complete discussion see 4.6, §4). For example, let $f(x,y)=x^2+y^4$, $g(x,y)=x^2-y^4$. The origin in $\mathbb{R}^2$ is the only critical point of $f,g$ and the associated quadratic form in both cases is $q(x,y)=x^2$, which is degenerate and positive semi-definite. However, $f$ has a strict minimum at $(0,0)$ while $g$ does not have a local extremum at $(0,0)$.

For a numerical algorithm to distinguish the three cases of Proposition 1.1, one examines the sequence $(\Gamma_k)_{1\leqslant k\leqslant n}=(\Gamma_1,\dots,\Gamma_n)$ of principal upper left $k\times k$ minors of the Hessian matrix $Hf(a)$ (Edwards [**6**, page 149]):

In the case $+,+,+,\dots$ of all positive signs, $q$ is positive definite; in the case $-,+,-,+,\dots$ of alternating signs, $q$ is negative definite; assuming $\det Hf(a)\neq 0$ (so that $q$ is nondegenerate), any other sequence of signs, or if zero terms occur, implies that $q$ attains both positive and negative values.

In the constrained case, $1\leqslant m<n$, recall (technical hypothesis) that the solution set of $g_1=g_2=\cdots=g_m=0$ is a submanifold of $\mathbb{R}^n$ near $a$. Let $q_L$ be the auxiliary quadratic form

$$q_L(h)=\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 L}{\partial x_i\partial x_j}(v)h_ih_j.$$

Subject to the constraints $g_\alpha=0, 1\leqslant\alpha\leqslant m$, the behaviour of $f$ near $a$ is analyzed by examining the restriction of the quadratic form $q_L$ to the linear subspace of $\mathbb{R}^n$ consisting of the tangent vectors to the manifold $A$ at $a$ (Edwards [**6**, page 154]). Thus, in analytic terms, one studies the quadratic form $q_L$ which is subject to the $m$ *linear* constraints:

$$\sum_{i=1}^{n}\frac{\partial g_\alpha}{\partial x_i}(a)h_i=0,\quad 1\leqslant\alpha\leqslant m. \tag{1.3}$$

PROPOSITION 1.2 (Edwards [**6**, page 154]). *Subject to the linear constraints* (1.3), *if $q_L$ is respectively positive definite, negative definite, attains both positive and negative values, then, subject to the constraints $g_\alpha=0, 1\leqslant\alpha\leqslant m$, the function $f$ has respectively, a strict local minimum, a strict local maximum, neither a local minimum nor a local maximum at $a$.*

REMARK 1.2. Proposition 1.1 is the special case of Proposition 1.2 obtained when $m=0$ ($L=f$; $q_L=q$).

REMARK 1.3. In the remaining cases, i.e., subject to the linear constraints (1.3), $q_L$ is degenerate and is positive or negative semi-definite, the character of the critical point $v$ is necessarily indeterminate at the level of second order derivatives (cf. (4.6), §4).

To complete the analysis of Proposition 1.2, one requires a numerical algorithm which classifies quadratic forms that are subject to linear constraints. H. B. Mann [**10**] classified positive or negative definite quadratic forms which are subject to linear constraints. The noted mathematical economist G. Debreu [**5**] generalized Mann's work by classifying positive or negative semi-definite quadratic forms which are subject to linear constraints. Since a quadratic form attains both positive and negative values if and only if it is neither positive nor negative semi-definite, Debreu's algorithms suffice to distinguish, in principle, the three cases of Proposition 1.2. This completes our review of sufficient conditions in the literature.

In this paper we propose a different approach to the problem of sufficient conditions for constrained local extrema. Our approach is to change coordinates in a neighbourhood of the point $a \in U$ so that the solution manifold $A$ is "flattened" near $a$ and so that the constraint functions, in new coordinates, are the simple constraint functions $g_\alpha(x_1, \dots, x_n) = x_\alpha, 1 \leqslant \alpha \leqslant m$, discussed above. In the presence of simple constraints, appropriate sufficient conditions are obtained in terms of an algorithm which classifies quadratic forms, without constraints.

Suppose $v = (c, a)$ is a critical point of the Lagrangian function $L$. A special feature of our approach is that the sufficient conditions are stated in terms of the $(m+n) \times (m+n)$ Hessian matrix $HL(v)$ of the function $L$ evaluated at the critical point $v$ (the matrix of second partial derivatives of $L$ evaluated at $v$).

$$\text{Explicitly, } HL(v) = \begin{bmatrix} 0 & \cdots & 0 & \dfrac{\partial g_1}{\partial x_1}(a) & \cdots & \dfrac{\partial g_1}{\partial x_n}(a) \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \dfrac{\partial g_m}{\partial x_1}(a) & \cdots & \dfrac{\partial g_m}{\partial x_n}(a) \\ \dfrac{\partial g_1}{\partial x_1}(a) & \cdots & \dfrac{\partial g_m}{\partial x_1}(a) & \dfrac{\partial^2 L}{\partial x_1^2}(v) & \cdots & \dfrac{\partial^2 L}{\partial x_1 \partial x_n}(v) \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{\partial g_1}{\partial x_n}(a) & \cdots & \dfrac{\partial g_m}{\partial x_n}(a) & \dfrac{\partial^2 L}{\partial x_n \partial x_1}(v) & \cdots & \dfrac{\partial^2 L}{\partial x_n^2}(v) \end{bmatrix}.$$

Note that the lower right $n \times n$ submatrix of $HL(v)$ is the Hessian matrix of the auxiliary quadratic form

$$q_L(h) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 L}{\partial x_i \partial x_j}(v) h_i h_j$$

which was employed in Proposition 1.2. A natural question concerns the precise statement of a second derivative test (i.e., a numerical algorithm) in terms of $HL(v)$ which distinguishes the three important cases:

(1.4) Subject to the constraints $g_\alpha = 0, 1 \leqslant \alpha \leqslant m$, $f$ has a (strict) local minimum value at $a$, or a (strict) local maximum value at $a$, or does not have a local extremum at $a$ (the latter is known as the saddle alternative).

The principal result, Theorem 1, distinguishes the three alternatives of (1.4) and in addition provides a complete saddle alternative, which seems to be new, in cases where $\det HL(v) = 0$. Theorem 1 employs a sequence of principal minors of $HL(v)$. This sequence is implicit in some older analyses of the problem (e.g., page 74 of Hancock [**9**]), but perhaps was first introduced

explicitly by H. B. Mann in this MONTHLY [**10**]. Unfortunately Mann's test, and its generalization by G. Debreu [**5**], have been largely forgotten in the mathematical literature. However, in recent years Mann's test has figured prominently in textbooks on the applications of mathematics to economics, where it is sometimes known as the *method of bordered Hessians* (Ostrosky and Koch [**13**], Y. Murata [**12**]; see also J. Marsden and A. J. Tromba [**11**]). A key observation underlying Theorem 1 is that Mann's test is easily understood in the case of simple constraints discussed above.

While the literature on the basic method of Lagrange multipliers is extensive, the literature on the second derivative test is sparse; in all the references of which I am aware, the saddle alternative of (1.4) is omitted (Hancock [**8**], [**9**], Carathéodory [**3**], J. Marsden and A. J. Tromba [**11**], Y. Murata [**12**], F. Bowman and F. A. Gerard [**2**], K. G. Binsmore [**1**]). Some texts on advanced calculus provide sufficient conditions for constrained local extrema but do not state a numerical algorithm for distinguishing the three alternatives of (1.4) (C. H. Edwards [**6**], R. Courant and F. John [**4**]).

We have attempted, in this paper, to develop a definitive form of the second derivative test for constrained local extrema. This has necessitated a more detailed study of the classification of quadratic forms than one finds in the usual literature. An appendix on quadratic forms has been added for this purpose.

**2. Statement of the theorem.** Let $(y_j)_{1 \leqslant j \leqslant r}$ be a non-trivial sequence of real numbers (i.e., not all the terms are zero). Let $y_k$ be the last non-zero term in this sequence.

DEFINITIONS:

2.1. The sequence $(y_j)$ is *positive semi-definite* if $y_j > 0$, $j = 1, \dots, k$.
2.2. The sequence $(y_j)$ is *negative semi-definite* if $(-1)^j y_j > 0$, $j = 1, \dots, k$.
2.3. The non-trivial sequence $(y_j)$ is of *saddle type* if neither (2.1) nor (2.2) holds. That is, either some $y_j = 0$, $j < k$, or the sequence of signs is not as indicated in (2.1), (2.2).
2.4. The sequence $(y_j)$ is *positive* (*negative*) *definite* if it is positive (negative) semi-definite and $k = r$ (i.e., the last term $y_r \neq 0$).

We interpret these definitions in Proposition 2.1 below.

*Notation*. (a) A principal $j$th order minor of a square $n \times n$ matrix $M$ is the determinant of a principal $j \times j$ submatrix of $M$ (i.e., a submatrix obtained by deleting $(n - j)$ rows and the corresponding $(n - j)$ columns of $M$).

(b) Let $M$ be a square $n \times n$ matrix. To each permutation $\pi$ of $\{1, \dots, n\}$ let $M(\pi)$ denote the matrix obtained from $M$ by permuting the rows and columns of $M$ by $\pi$. Associated to $\pi$ is the linear change of coordinates: $(x_1, \dots, x_n) \to (x_{\pi(1)}, \dots, x_{\pi(n)})$. Thus $M(\pi) = E^{-1} \times M \times E$, for some orthogonal matrix $E$ (depending on $\pi$).

The principal background algebraic result that we require is as follows:

PROPOSITION 2.1. *Let $Q$ be a real-valued quadratic form on $\mathbb{R}^r$ whose associated matrix is $M$ (in some basis of $\mathbb{R}^r$).*

(a) *$Q$ is positive definite (negative definite) if and only if the sequence $(y_j)_{1 \leqslant j \leqslant r}$ of principal upper left minors of $M$ is positive definite (negative definite).*

(b) *$Q$ attains both positive and negative values if and only if there is a permutation $\pi$ of $\{1, \dots, r\}$ such that the sequence $(y_j(\pi))_{1 \leqslant j \leqslant r}$ of principal upper left minors of $M(\pi)$ is non-trivial and of saddle type.*

*Proof*. Conclusion (a) is well known (Edwards [**6**, page 149]). Conclusion (b) does not seem to appear in the literature. The "if" part of (b) is trivial and is a consequence of Lemma 2.1 below (note that $M(\pi)$ is the matrix for $Q$ obtained by permuting the given basis of $\mathbb{R}^r$ by $\pi$). The "only if" part is proved in the Appendix.

LEMMA 2.1. *Let $q$ be a real-valued quadratic form on $\mathbb{R}^n$ whose associated matrix is $N$ (in some basis of $\mathbb{R}^n$). If the sequence $(y_j)_{1 \leqslant j \leqslant n}$ of principal upper left minors of $N$ is non-trivial and of saddle type, then $q$ attains both positive and negative values.*

*Proof.* Suppose $y_k$ is the last non-zero term in the sequence. Let $V$ be the subspace of $\mathbb{R}^n$ generated by the first $k$ vectors in the given basis of $\mathbb{R}^n$. Denote by $\tilde{q}$ the quadratic form on $V$ obtained by restriction: $\tilde{q}(x) = q(x)$, all $x \in V$. Since $y_k \neq 0$, it follows that $\tilde{q}$ is non-degenerate. Since the sequence $(y_j)_{1 \leqslant j \leqslant k}$ is of saddle type, it follows from the classification of non-degenerate quadratic forms (Edwards [**6**, page 149]) that $\tilde{q}$ (and hence $q$) attains both positive and negative values.

With these preliminaries, we state our second derivative test as follows: Let $\Gamma_k = (-1)^m \times$ principal upper left $k$th order minor of the Hessian Matrix $HL(v)$, $1 \leqslant k \leqslant m + n$. In particular, $\Gamma_{n+m} = (-1)^m \det HL(v)$. Note that in case $m = 0$ (no constraints), then $L = f$ and $v = a \in U$ is a critical point of $f$. In this case, $\Gamma_k$ is the principal upper left $k$th order minor of the Hessian matrix $Hf(a)$, $1 \leqslant k \leqslant n$. Let $\pi$ be a permutation of $\{1, \ldots, m+n\}$; let $\Gamma_k(\pi) = (-1)^m \times$ principal upper left $k$th order minor of the matrix $HL(v)(\pi)$, $1 \leqslant k \leqslant m + n$. Only the sequences

$$\left(\Gamma_{2m+p}\right)_{1 \leqslant p \leqslant n-m} = \left(\Gamma_{2m+1}, \ldots, \Gamma_{m+n}\right) \quad \text{and} \quad \left(\Gamma_{2m+p}(\pi)\right)_{1 \leqslant p \leqslant n-m}$$

are used in the test.

THEOREM 1 (Second Derivative Test). *Suppose $v = (c, a) \in \mathbb{R}^m \times U$ is a critical point of the Lagrangian function $L$. Assume $f, g_\alpha : U \to \mathbb{R}$, $1 \leqslant \alpha \leqslant m$; $m < n$, are of differentiability class $C^3$ and that $g = (g_1, \ldots, g_m) : U \to \mathbb{R}^m$ has maximal rank $(= m)$ at $a$. In particular, by relabelling the variables $x_1, \ldots, x_n$ if necessary, assume*

$$\frac{\partial(g_1, \ldots, g_m)}{\partial(x_1, \ldots, x_m)}(a) \neq 0.$$

(a) *Suppose $\Gamma_{m+n} \neq 0$ (i.e.,* $\det HL(v) \neq 0$).

(i) *If the sequence $(\Gamma_{2m+p})_{1 \leqslant p \leqslant n-m}$ is positive definite, then subject to the constraints, $f$ has a strict local minimum at $a \in U$.*

(ii) *If the sequence $(\Gamma_{2m+p})_{1 \leqslant p \leqslant n-m}$ is negative definite, then subject to the constraints, $f$ has a strict local maximum at $a \in U$.*

(b) *Suppose the sequence $(\Gamma_{2m+p})_{1 \leqslant p \leqslant n-m}$ is nontrivial and of saddle type (in particular we do not assume* $\det HL(v) \neq 0$, *only that $\Gamma_{2m+p} \neq 0$ for some $p$). Then subject to the constraints, $f$ has neither a local maximum nor a local minimum at $a \in U$. This case is known as the saddle alternative.*

(b)′ (*Comprehensive saddle alternative*). *Suppose there is a permutation $\pi$ of the last $(n - m)$ rows and columns of $HL(v)$ (this corresponds to the permutation of the last $(n - m)$ variables in $\mathbb{R}^n : (x_1, \ldots, x_n) \to (x_1, \ldots, x_m, x_{\pi(m+1)}, \ldots, x_{\pi(n)})$) such that the sequence $(\Gamma_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}$ associated to the matrix $HL(v)(\pi)$ is non-trivial and of saddle type. Then subject to the constraints, $f$ has neither a local maximum nor a local minimum at $a \in U$.*

REMARKS.

2.1. Suppose, in (b), $\Gamma_{m+n} \neq 0$ (i.e., $\det HL(v) \neq 0$). This case is known as the non-degenerate saddle alternative and is equivalent to: not ((i) or (ii)). Thus if $\Gamma_{m+n} \neq 0$, the second derivative test is complete. Also, if $\det HL(v) \neq 0$, then (b), (b)′ are equivalent.

2.2. Clearly (b) is a special case of (b)′ (let $\pi = id$). In practice, if (a) fails, then (b) usually applies. For purposes of illustration, Example 3 below demonstrates that (b)′ may apply where (b) fails, in case $\det \mathrm{HL}(v) = 0$.

2.3. The only indeterminate case—in which Theorem 1 implies no conclusion about the character

of the critical point $v$—is that for which $\Gamma_{m+n} = \det HL(v) = 0$ and (b)′ fails. In terms of the analysis of sufficient conditions provided by Proposition 1.2, §1, this corresponds to the case that, subject to the linear constraints (1.3), the auxiliary quadratic form $q_L$ is degenerate and is positive or negative semi-definite (cf. (4.5), (4.6), §4 for complete details). As will be explained in §4, the degenerate semi-definite cases are necessarily indeterminate at the level of second order partial derivatives. In this sense, the statement of Theorem 1 cannot be improved.

2.4. In case $m = 0$ (i.e., no constraints) and $\Gamma_n = \det Hf(a) \neq 0$, Theorem 1 is the well-known second derivative test for a local extremum of a function $f$ at a critical point (Edwards [**6**, page 145]). Note that conclusions (b), (b)′ improve and complete this classical second derivative test when $\det Hf(a) = 0$.

2.5. Proofs of part (a) of Theorem 1 with $n = 3$ and $1 \leqslant m \leqslant 2$ appear in various textbooks (e.g., Bowman and Gerard [**2**]). However, these textbook treatments generally either ignore the saddle alternative or assume that $\det HL(v) \neq 0$.

2.6. In case $m = n - 1$, there is only one term in the sequence, i.e., $\Gamma_{m+n} = (-1)^m \det HL(v)$. In this case, the hypothesis that $g: U \to \mathbb{R}^m$ has rank $m$ at $a \in U$ is sufficient (i.e., no relabelling of the variables is necessary). Note also there is no saddle alternative in this case.

We illustrate Theorem 1 with some examples.

EXAMPLE 1. $f(x, y, z) = xyz$ subject to $x^2 + y^2 + z^2 - 1 = 0$. This example is taken from Marsden and Tromba [**11**], page 229. A critical point for

$$L = xyz + \lambda(x^2 + y^2 + z^2 - 1)$$

is $v = (0,1,0,0)$. For this critical point, one calculates $\Gamma_3 = 0$; $\Gamma_4 = -\det HL(v) = -4$. Applying Theorem 1, we see that this is the saddle case. Lacking a saddle alternative, Marsden and Tromba (uncritically) conclude that $v$ falls in the test fails, i.e., indeterminate case.

EXAMPLE 2. To find the minimum distance between a point $(x_1, x_3)$ of the circle $x^2 + y^2 = 2$ and a point $(x_2, x_4)$ of the line $x + y = 4$ (see page 156 of Edwards [**6**]) we seek a local minimum value of

$$f(x_1, x_2, x_3, x_4) = (x_1 - x_2)^2 + (x_3 - x_4)^2$$

subject to the two constraints

$$g_1(x_1, x_2, x_3, x_4) = x_1^2 + x_3^2 - 2 = 0;$$
$$g_2(x_1, x_2, x_3, x_4) = x_2 + x_4 - 4 = 0.$$

The critical points of the Lagrangian $L = f + \lambda_1 g_1 + \lambda_2 g_2$ are: $v_1 = (1, -2, 1, 2, 1, 2)$ and $v_2 = (-3, -6, -1, 2, -1, 2)$. From the bordered Hessian

$$HL(v) = \begin{bmatrix} 0 & 0 & 2x_1 & 0 & 2x_3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 2x_1 & 0 & 2 + 2\lambda_1 & -2 & 0 & 0 \\ 0 & 1 & -2 & 2 & 0 & 0 \\ 2x_3 & 0 & 0 & 0 & 2 + 2\lambda_1 & -2 \\ 0 & 1 & 0 & 0 & -2 & 2 \end{bmatrix},$$

one calculates that $\Gamma_5 = 32$ and $\Gamma_6 = 64$ at $v_1$, while $\Gamma_5 = -32, \Gamma_6 = -192$ at $v_2$. Applying Theorem 1, one concludes that $f$ has a local minimum (in fact an absolute minimum) at $(1,2,1,2)$, corresponding to the points $(1,1)$ on the circle and $(2,2)$ on the line, while $f$ does not have a local extremum at $(-1,2,-1,2)$, corresponding to the points $(-1,-1)$ on the circle and $(2,2)$ on the line. These conclusions are clear geometrically.

EXAMPLE 3. $f(x, y, z, t) = y^2 + z^2 - t^2 - x^2$, subject to $x - z = 0$. Clearly $f$ is a non-degenerate quadratic form which becomes degenerate and attains both positive and negative values on the three dimensional subspace defined by the linear equation $x = z$ on $\mathbb{R}^4$. We check this by Theorem 1. The critical points of

$$L = y^2 + z^2 - t^2 - x^2 + \lambda(x - z)$$

are of the form $v = (2t, t, 0, t, 0), t \in \mathbb{R}$. Computing the $5 \times 5$ Hessian of $L$, one calculates $\Gamma_3 = 2; \Gamma_4 = 0; \Gamma_5 = 0$. The sequence $2, 0, 0$ is positive semi-definite so Theorem 1, part (b) does not apply. However, by interchanging $z$, $t$ one obtains the Lagrangian

$$y^2 + t^2 - z^2 - x^2 + \lambda(x - t)$$

for which one computes $\Gamma_3 = 2; \Gamma_4 = -4; \Gamma_5 = 0$. Since the sequence $2, -4, 0$ is of saddle type, one concludes from Theorem 1, part (b)$'$, that subject to the constraint $x = z$, $f$ does not have a local extremum at any point $(t, 0, t, 0) \in \mathbb{R}^4, t \in \mathbb{R}$.

**3.** *Proof of Theorem* 1. We first prove the theorem for the simple constraints $x_\alpha = 0$, $1 \leqslant \alpha \leqslant m$, introduced in §1. The analysis of $HL(v)$ for these constraints will elucidate the role of the signed minors $\Gamma_{2m+p}, 1 \leqslant p \leqslant n - m$. Let

$$\tilde{f}(x_{m+1}, \ldots, x_n) = f(0, \ldots, 0, x_{m+1}, \ldots, x_n).$$

The associated Lagrangian function is $L = f + \sum_{\alpha=1}^{m} \lambda_\alpha x_\alpha$, whose critical points are of the form

$$v = (c, a) \in \mathbb{R}^m \times U,$$

where

(i) $a = (0, \ldots, 0, a_{m+1}, \ldots, a_n) \in \mathbb{R}^n$ and $\tilde{a} = (a_{m+1}, \ldots, a_n)$ is a critical point of $\tilde{f}$.

(ii) $c = \left( -\dfrac{\partial f}{\partial x_1}(a), \ldots, -\dfrac{\partial f}{\partial x_m}(a) \right) \in \mathbb{R}^m$.

(3.1) Evidently,

$$\frac{\partial^2 L}{\partial x_i \partial x_j}(v) = \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j}(\tilde{a}), \quad m + 1 \leqslant i, \quad j \leqslant n.$$

Consequently, $HL(v)$ assumes the block matrix form

$$HL(v) = \begin{bmatrix} 0_{m,m} & I_m & 0_{m,n-m} \\ I_m & B & C \\ 0_{n-m,m} & C^T & H\tilde{f}(\tilde{a}) \end{bmatrix},$$

where $H\tilde{f}(\tilde{a}) = \left[ \dfrac{\partial^2 \tilde{f}}{\partial x_i \partial x_j}(\tilde{a}) \right]$ is the Hessian matrix of $\tilde{f}$ evaluated at the critical point $\tilde{a} = (a_{m+1}, \ldots, a_n)$; $0_{r,s}$ is the zero $r \times s$ matrix, and $I_r$ the identity $r \times r$ matrix. Let $(\Delta_p)_{1 \leqslant p \leqslant n-m}$ be the sequence of principal upper left minors of the Hessian $H\tilde{f}(\tilde{a})$; let $(S_k)_{1 \leqslant k \leqslant n+m}$ be the sequence of principal upper left minors of the Hessian $HL(v)$. By definition,

$$\Gamma_k = (-1)^m S_k, \qquad 1 \leqslant k \leqslant m + n.$$

Employing elementary properties of determinants, one verifies easily that

$$(3.2) \qquad \text{(i)} \quad S_{2m} = (-1)^m \qquad \text{(ii)} \quad S_{2m+p} = (-1)^m \Delta_p, \quad 1 \leqslant p \leqslant n - m.$$

(3.3) Consequently, $\Gamma_{2m+p} = (-1)^{2m} \Delta_p = \Delta_p, 1 \leqslant p \leqslant n - m$. That is, the sequences $(\Gamma_{2m+p})_{1 \leqslant p \leqslant n-m}$ and $(\Delta_p)_{1 \leqslant p \leqslant n-m}$ coincide.

The relations (3.3) explain the sequence of signed determinants introduced by Mann **[10]**. To prove the theorem, we observe that:

(3.4) Subject to the constraints $x_\alpha = 0, 1 \leqslant \alpha \leqslant m$, $f$ has a (strict) local minimum or a (strict) local maximum or does not have a local extremum at $a \in A$, if and only if the function $\tilde{f}$ has respectively a (strict) local minimum or a (strict) local maximum or does not have a local extremum at $\tilde{a} = (a_{m+1}, \ldots, a_n)$.

Let

$$\tilde{q}(h) = \frac{1}{2} \sum_{i,j=m+1}^{n} \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j}(\tilde{a}) h_i h_j$$

be the quadratic form associated to the function $\tilde{f}$ at the critical point $\tilde{a}$ (cf. (1.2), §1). Applying Proposition 1.1 to $\tilde{f}$ and also applying the classification results Proposition 2.1 (a), Lemma 2.1 to the quadratic form $\tilde{q}$, we find, from the relations (3.3), (3.4) above, that conclusions (a), (b) of Theorem 1 are proved.

It remains to prove conclusion (b)′. From the displayed form above of the matrix $HL(v)$, it is evident that, for any permutation $\pi$ of the last $(n-m)$ rows and columns of $HL(v)$, the principal upper left $2m \times 2m$ submatrix of $HL(v)$ remains unchanged, while $H\tilde{f}(\tilde{a})$ is changed to $H\tilde{f}(\tilde{a})(\pi)$.

(3.5) Consequently, the proof of (3.3) shows that the sequences $(\Gamma_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}$ and $(\Delta_p(\pi))_{1 \leqslant p \leqslant n-m}$ coincide.

Applying Proposition 1.1 to $\tilde{f}$ and also Proposition 2.1 (b) to the quadratic form $\tilde{q}$, we find from the relations (3.4), (3.5) above, that conclusion (b)′ of Theorem 1 is proved.

To prove Theorem 1 for general constraints, we change coordinates to reduce to the case of simple constraints. Let

$$\varphi : \mathbb{R}^m \times U \to \mathbb{R}^m \times \mathbb{R}^n = \mathbb{R}^{m+n}$$

be the map

$$\varphi(\lambda_1, \ldots, \lambda_m, x_1, \ldots, x_n) = (\lambda_1, \ldots, \lambda_m, z_1, \ldots, z_n),$$

where

$$z_\alpha = g_\alpha(x_1, \ldots, x_n), \quad 1 \leqslant \alpha \leqslant m; z_{m+i} = x_{m+i}, \quad m+1 \leqslant i \leqslant n.$$

In block matrix notation,

$$D\varphi(v) = \begin{bmatrix} I_m & 0_{m,n} \\ 0_{m,m} & Dg(a) \\ 0_{n-m,2m} & I_{n-m}4 \end{bmatrix}. \tag{3.6}$$

Since $\dfrac{\partial(g_1, \ldots, g_m)}{\partial(x_1, \ldots, x_m)}(a) \neq 0$, it is clear that the $C^3$ map $\varphi$ has maximal rank $(= m+n)$ at $v = (c, a) \in \mathbb{R}^m \times U$ and hence, by the Inverse Function Theorem [**11**], $(\lambda_1, \ldots, \lambda_m, z_1, \ldots, z_n)$ defines a coordinate system in a neighbourhood of $w = \varphi(v)$ in $\mathbb{R}^{m+n}$. Write

$$\varphi^{-1}(\lambda, z) = (\lambda, \psi(z)) \in \mathbb{R}^m \times U,$$

and let

$$\tilde{L} = L \circ \varphi^{-1} = f(\psi(z)) + \sum_{\alpha=1}^{m} \lambda_\alpha z_\alpha.$$

Thus $\tilde{L}$ is the Lagrangian associated to the $C^3$ function $f \circ \psi$, subject to the constraints $z_\alpha = 0$, $1 \leqslant \alpha \leqslant m$. Since $v$ is a critical point of $L$, then $w = \varphi(v)$ is a critical point of the Lagrangian $\tilde{L} = L \circ \varphi^{-1}$. Evidently, the following equivalence obtains:

(3.7) Subject to the constraints $g_\alpha = 0, 1 \leqslant \alpha \leqslant m$, $f$ has a (strict) local minimum or a (strict) local maximum or $f$ does not have a local extremum at $a \in A$ if and only if, subject to the *simple* constraints $z_\alpha = 0, 1 \leqslant \alpha \leqslant m, f \circ \psi$ has respectively a (strict) local minimum or a (strict) local

maximum or $f \circ \psi$ does not have a local extremum at $a_0, \psi(a_0) = a$.

The equivalence (3.7) indicates that we should compare the sequences of principal upper left minors of the Hessians $HL(v), H\tilde{L}(w)$. This is accomplished by the following lemma.

LEMMA 3.1. *$HL(v) = P^T \times H\tilde{L}(w) \times P$, where $P = D\varphi(v)$.*

COROLLARY 3.1.1. *Let $\pi$ be a permutation of $\{1,\dots,m+n\}$. Then $HL(v)(\pi) = P^T(\pi) \times HL(w)(\pi) \times P(\pi)$.*

COROLLARY 3.1.2. *Suppose, in addition, $\pi$ fixes pointwise $\{1,\dots,2m\}$ (i.e., $\pi$ induces a permutation of the last $(n-m)$ rows and columns of $(m+n)\times(m+n)$ matrices). Let $(\tilde{\Gamma}_k(\pi))_{1 \leqslant k \leqslant m+n}$ be the sequence of principal upper left minors of $H\tilde{L}(w)(\pi)$. Then*

$$\Gamma_{2m+p}(\pi) = \left[\frac{\partial(g_1,\dots,g_m)}{\partial(x_1,\dots,x_m)}(a)\right]^2 \cdot \tilde{\Gamma}_{2m+p}(\pi), \qquad 0 \leqslant p \leqslant n-m.$$

COROLLARY 3.1.3. *For each permutation $\pi$ as in Corollary 3.1.2, both the sequences $(\Gamma_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}, (\tilde{\Gamma}_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}$, are, respectively, trivial or positive (semi-) definite or negative (semi-) definite or of saddle type.*

It is clear that Corollary 3.1.3 and the equivalence (3.7) above prove Theorem 1 for general constraints.

*Proof of Lemma* 3.1. This lemma is a special case of a well-known result: A change of coordinates induces a congruence of the Hessians at the corresponding critical points. Explicitly, recall that $L = \tilde{L} \circ \varphi$. Introducing the auxiliary notation,

$$(u_1,\dots,u_{m+n}) = (\lambda_1,\dots,\lambda_m, x_1,\dots,x_n),$$
$$(v_1,\dots,v_{m+n}) = (\lambda_1,\dots,\lambda_m, z_1,\dots,z_n),$$

and

$$\varphi = (\varphi_1,\dots,\varphi_{m+n}) : \mathbb{R}^m \times U \to \mathbb{R}^{m+n},$$

we find, from a straightforward Chain-Rule computation, that

$$\frac{\partial^2 L}{\partial u_i \partial u_j}(v) = \sum_{r,s=1}^{m+n} \frac{\partial^2 \tilde{L}}{\partial v_r \partial v_s}(w) \frac{\partial \varphi_r}{\partial u_i}(v) \frac{\partial \varphi_s}{\partial u_j}(v), \quad 1 \leqslant i, \quad j \leqslant m+n,$$

at the critical points $v$ and $w = \varphi(v)$. That is,

$$HL(v) = P^T \times H\tilde{L}(w) \times P, \quad \text{where } P = D\varphi(v).$$

*Proof of Corollary* 3.1.1. Let $E$ be the orthogonal $(m+n)\times(m+n)$ matrix corresponding to the permutation $\pi$. Thus

$$\begin{aligned} HL(v)(\pi) &= E^{-1} \times HL(v) \times E = E^{-1} \times P^T \times H\tilde{L}(w) \times P \times E \\ &= P^T(\pi) \times HL(w)(\pi) \times P(\pi). \end{aligned}$$

*Proof of Corollary* 3.1.2. From (3.6), the principal upper left $2m \times 2m$ submatrix of $P$ is

$$P_{2m} = \begin{bmatrix} I_m & 0_{m,m} \\ 0_{m,n} & D_1(g)(a) \end{bmatrix},$$

where

$$D_1(g)(a) = \left[\frac{\partial g_i}{\partial x_j}(a)\right]_{1 \leqslant i,j \leqslant m}.$$

In particular,

$$\det P_{2m} = \frac{\partial(g_1, \ldots, g_m)}{\partial(x_1, \ldots, x_m)}(a)(\neq 0).$$

Observe that if $\pi$ is *any* permutation of the last $(n-m)$ rows and columns of $P$, the matrix $P(\pi)$ assumes the form

$$P(\pi) = \begin{bmatrix} P_{2m} & B \\ 0_{n-m,2m} & I_{n-m} \end{bmatrix}, \tag{3.8}$$

where $B$ is a $2m \times (n-m)$ matrix which depends on $\pi$.

Since $HL(v)(\pi) = P^T(\pi) \times H\tilde{L}(w)(\pi) \times P(\pi)$, a simple block matrix calculation, employing (3.8) (note also $P^T(\pi) = P(\pi)^T$), proves Corollary 3.1.2. This completes the proof of Theorem 1 for general constraints.

**4.** In this section we explain the connection between Theorem 1 and the classification result Proposition 1.2, §1. To begin, we classify quadratic forms which are subject to linear constraints.

Let $q = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}x_ix_j$ be a real valued quadratic form on $\mathbb{R}^n$; $A = [a_{ij}]_{1 \leqslant i,j \leqslant n}$. Suppose $q$ is subject to $m$ linear constraints, $\sum_{j=1}^{n} b_{ij}x_j = 0$, $1 \leqslant i \leqslant m$; $m < n$. Let $B = [b_{ij}]_{1 \leqslant i \leqslant m; 1 \leqslant j \leqslant n}$ be the coefficient matrix of the constraints. We assume the $m \times m$ submatrix $[b_{ij}]_{1 \leqslant i,j \leqslant m}$ is non-singular. Evidently $v = (0,0) \in \mathbb{R}^m \times \mathbb{R}^n$ is a critical point of the Lagrangian function

$$L = q + \sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij}x_j\lambda_i.$$

Note that the Hessian matrix $HL(v)$ is:

$$HL(v) = \begin{bmatrix} 0_{m,m} & B \\ B^T & A \end{bmatrix}. \tag{4.1}$$

Recall, §2, the sequences $(\Gamma_k)_{1 \leqslant k \leqslant m+n}$, $(\Gamma_k(\pi))_{1 \leqslant k \leqslant m+n}$ of principal upper left minors respectively of the matrices $HL(v)$, $HL(v)(\pi)$.

THEOREM 2. *Subject to the above linear constraints,*

(a) *$q$ is positive (negative) definite if and only if the sequence $(\Gamma_{2m+p})_{1 \leqslant p \leqslant n-m}$ is positive (negative) definite.*

(b) *$q$ attains both positive and negative values if and only if there is a permutation $\pi$ of the last $(n-m)$ variables in $\mathbb{R}^n$ (i.e., $(x_1, \ldots, x_n) \to (x_1, \ldots, x_m, x_{\pi(m+1)}, \ldots, x_{\pi(n)})$) such that the sequence $(\Gamma_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}$ is non-trivial and of saddle type.*

(c) *$q$ is positive (negative) semi-definite if and only if for each permutation $\pi$ of the last $(n-m)$ variables of $\mathbb{R}^n$ the sequence $(\Gamma_{2m+p}(\pi))_{1 \leqslant p \leqslant n-m}$ is trivial or positive (negative) semi-definite.*

*Proof.* To prove Theorem 2, one essentially applies Theorem 1 to quadratic forms which are subject to linear constraints. Suppose first of all that the linear constraints are simple: $x_i = 0$, $1 \leqslant i \leqslant m$. Let $\tilde{q}(x_{m+1}, \ldots, x_n)$ be the quadratic form

$$q(0, \ldots, 0, x_{m+1}, \ldots, x_n) = \frac{1}{2} \sum_{i,j=m+1}^{n} a_{ij}x_ix_j.$$

In the case of simple constraints, Theorem 2 follows from the relations (3.3), (3.5), §3, (let $f = q$; $\tilde{f} = \tilde{q}$ in Theorem 1) and the classification results applied to $\tilde{q}$ provided by Proposition 2.1, §2, and Proposition A.1 in the Appendix. For general linear constraints, note that in Theorem 1, if $f = q$ and

$$g_\alpha(x_1, \ldots, x_n) = \sum_{j=1}^{n} b_{\alpha j}x_j, \qquad 1 \leqslant \alpha \leqslant m,$$

then the change of coordinates map, $\varphi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{m+n}$, constructed in Theorem 1 is a *linear* map. Consequently, in new coordinates, the function $q \circ \psi$ is a quadratic form which is subject to simple constraints. Thus Theorem 2 follows from Corollary 3.1.3, §3, and the classification, proved above, for quadratic forms which are subject to simple constraints.

REMARK 4.1. Conclusion (a) was first established by H. B. Mann [**10**]. Conclusion (c) sharpens and simplifies the classification scheme obtained by G. Debreu [**5**] for semi-definite quadratic forms. (In particular Debreu's results are stated in terms of *all* permutations of the variables in $\mathbb{R}^n$.) Despite its obvious utility for the theory of Lagrange multipliers (cf. Propositions 1.1, 1.2, §1), conclusion (b) of Theorem 2 appears to be new.

Turning now to the study of the indeterminate case for Theorem 1, recall that Proposition 1.2, §1, is stated in terms of the auxiliary quadratic form on $\mathbb{R}^n$,

$$q_L(h) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 L}{\partial x_i \partial x_j}(v) h_i h_j,$$

which is subject to the $m$ linear constraints (1.3):

$$\sum_{i=1}^{n} \frac{\partial g_\alpha}{\partial x_i}(a) h_i = 0, \qquad 1 \leqslant \alpha \leqslant m.$$

(4.2) OBSERVATION. Let $L_1$ be the Lagrangian function associated to the quadratic form $q_L$ and the linear constraints (1.3). It is clear from (4.1) that the Hessian matrix $HL_1(0)$ is exactly the Hessian matrix $HL(v)$ defined in §1.

Applying (4.2) and Theorem 2, we establish the following results:

(4.3) Conclusion (a) of Theorem 1 obtains, if and only if, subject to (1.3), the quadratic form $q_L$ is positive or negative definite.

(4.4) Conclusion (b)′ of Theorem 1 obtains, if and only if, subject to the constraints (1.3), $q_L$ attains both positive and negative values.

(4.5) The only indeterminate case for Theorem 1, at the level of second order derivatives, occurs if and only if $\det HL(v) = 0$ and (b)′ fails, i.e., subject to the linear constraints (1.3), the auxiliary quadratic form $q_L$ is degenerate and is positive or negative semi-definite, a numerical algorithm for which is provided by Theorem 2 (c).

(4.6) For completeness, we show that this indeterminate case is necessarily indeterminate at the level of second order derivatives.

Let $V$ be the $(n-m)$-dimensional linear subspace of $\mathbb{R}^n$ defined by the linear constraints (1.3). There is a basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ such that $(e_1, \ldots, e_{n-m})$ is a basis of $V$ and, with respect to coordinates $(y_1, \ldots, y_n)$ in this basis, the quadratic form $q_L$ is diagonalized:

$$q_L(y_1, \ldots, y_n) = \varepsilon(y_1^2 + \cdots + y_k^2), \quad \text{where} \quad \varepsilon = 1\,(\varepsilon = -1),$$

if $q_L$ is positive (negative) semi-definite on $V$ and $k$ is the maximal dimension of a subspace of $V$ on which $q_L$ is positive (negative) definite. Since $q_L$ is degenerate on $V$, then $k < n - m$ ($q_L = 0$ if $k = 0$). In the basis above (assuming $m > 0$), the linear constraints (1.3) are equivalent to the simple constraints: $y_i = 0, \quad i = n - m + 1, \ldots, n$. Let

$$f(y_1, \ldots, y_n) = q_L(y_1, \ldots, y_n) + y_{k+1}^4, \quad g(y_1, \ldots, y_n) = q_L(y_1, \ldots, y_n) - y_{k+1}^4,$$

subject to the simple constraints (if $m > 0$) $y_i = 0, \quad i = n - m + 1, \ldots, n$. Subject to these simple constraints, if $\varepsilon = 1$, then $f$ has an absolute minimum at the origin in $\mathbb{R}^n$, while $g$ does not have a local extremum at the origin in $\mathbb{R}^n$, although the auxiliary quadratic form is $q_L$ in both cases. Similar conclusions obtain if $\varepsilon = -1$. Consequently, the indeterminate case (4.5) is necessarily indeterminate.

REMARK 4.2. It follows from (4.3), (4.4) above that Theorem 1 provides a numerical algorithm which distinguishes the three cases of Proposition 1.2, §1.

This completes our study of Theorem 1 and of the problem of constrained local extrema.

APPENDIX. In this appendix we recall some results from linear algebra and prove Proposition 2.1 (b), §2. Let $Q$ be a real-valued quadratic form on $\mathbb{R}^r$ whose associated matrix is $M$ (in some basis of $\mathbb{R}^r$). For each permutation $\pi$ of $\{1,2,\ldots,r\}$, recall the sequence $(y_j(\pi))_{1\leqslant j\leqslant r}$ of principal upper left minors of the matrix $M(\pi)$ (cf. §2 for this notation). Let

$$(x,y)=\frac{1}{2}(Q(x+y)-Q(x)-Q(y))$$

be the associated bilinear form.

PROPOSITION A.1. (a) *$Q$ is positive semi-definite if and only if, for each permutation $\pi$, the sequence $(y_j(\pi))$ is either trivial or positive semi-definite.*

(b) *$Q$ is negative semi-definite if and only if for each permutation $\pi$, the sequence $(y_j(\pi))$ is either trivial or negative semi-definite.*

*Proof.* It is enough to prove (a). To prove (b) one considers the quadratic form $-Q$. Employing the classical classification of semi-definite quadratic forms (Debreu [5], Gantmacher [7], page 307) $Q$ is positive semi-definite if and only if each principal $j$th order minor of $M$ is $\geqslant 0$, $j=1,2,\ldots,r$. Note that each principal $j$th order minor of $M$ is the principal upper left $j$th order minor of $M(\pi)$, for some permutation $\pi$. Thus, one obtains the condition: $y_j(\pi)\geqslant 0$, $j=1,\ldots,r$, for all permutations $\pi$. However, if the sequence $(y_j(\pi))_{1\leqslant j\leqslant r}$ is non-trivial, then *only* a sequence of $+$ signs followed by a sequence of zeros can occur. Indeed, any other non-trivial sequence of $+$ signs and zeros would constitute a sequence of saddle type and hence it would follow, by Lemma 2.1, §2, that $Q$ is not positive semi-definite. Thus Proposition A.1 sharpens the classical classification of semi-definite quadratic forms.

LEMMA A.1. *Suppose there are permutations $\pi$, $\sigma$ of $\{1,2,\ldots,r\}$ such that the sequences $(y_j(\pi))_{1\leqslant j\leqslant r}$ and $(y_j(\sigma))_{1\leqslant j\leqslant r}$ are nontrivial and positive semi-definite, respectively, negative semi-definite. Then there is a permutation $\tau$ of $\{1,2,\ldots,r\}$ such that the sequence $(y_j(\tau))_{1\leqslant j\leqslant r}$ is non-trivial and of saddle type.*

*Proof.* Let $y_1(\pi)=Q(e)>0$; $y_1(\sigma)=Q(f)<0$, where $e,f$ are elements of the given basis of $\mathbb{R}^r$. Let $\tau$ be a permutation of $\{1,2,\ldots,r\}$ such that the principal upper left $2\times 2$ submatrix of $M(\tau)$ is $\begin{bmatrix}(f,f) & (e,f)\\ (e,f) & (e,e)\end{bmatrix}$. Evidently,

$$y_1(\tau)=Q(f)<0,\quad y_2(\tau)=Q(f)Q(e)-[(e,f)]^2<0.$$

Consequently the sequence $(y_j(\tau))_{1\leqslant j\leqslant r}$ is non-trivial and of saddle type.

*Proof of Proposition* 2.1 (b). Suppose $Q$ attains both positive and negative values. Then $Q$ is neither positive nor negative semi-definite. By Proposition A.1, either there is a permutation $\pi$ for which the sequence $(y_j(\pi))_{1\leqslant j\leqslant r}$ is non-trivial and of saddle type, or there are permutations $\pi$, $\sigma$ for which the hypothesis of Lemma A.1 holds. In either case there is a permutation $\gamma$ such that the sequence $(y_j(\gamma))_{1\leqslant j\leqslant r}$ is non-trivial and of saddle type. Since Lemma 2.1, §2 proves the "if" part of (c), this completes the proof of Proposition 2.1 (b).

## References


1. K. G. Binmore, Calculus, Cambridge University Press, 1983.
2. F. Bowman and F. A. Gerard, Higher Calculus, Cambridge University Press, 1967.
3. C. Carathéodory, Calculus of Variations and Partial Differential Equations II, Holden-Day, 1965.
4. R. Courant and F. John, Introduction to Calculus and Analysis, Interscience Publishers, 1965.
5. G. Debreu, Definite and semidefinite quadratic forms, Econometrica, 20 (1952) 295–300.

6. C. H. Edwards, Jr., Advanced Calculus of Several Variables, Academic Press, 1973.
7. F. R. Gantmacher, The Theory of Matrices, Vol. I, Chelsea, New York, 1977.
8. H. Hancock, Theory of Maxima and Minima, Ginn, 1917.
9. ______, Lectures on the Theory of Maxima and Minima of Functions of Several Variables, University of Cincinnati Bulletin No. 13, Series II, Vol. II.
10. H. B. Mann, Quadratic forms with linear constraints, this MONTHLY, 50 (1943) 430–433.
11. J. Marsden and A. J. Tromba, Vector Calculus, 2nd ed., Freeman, 1976.
12. Y. Murata, Mathematics for Stability and Optimization of Economic Systems, Academic Press, 1977.
13. A. L. Ostrosky and J. V. Koch, Introduction to Mathematical Economics, Houghton Mifflin, 1979.
14. Ch. J. de la Vallée Poussin, Cours d'Analyse Infinitésimale, Dover Publications, New York, Vol. 1, 8th ed., Chapter III, Section 4 and Chapter IV, Section 3, 1946.

# THE SEARCH FOR A ROLLE'S THEOREM IN THE COMPLEX DOMAIN

MORRIS MARDEN
*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201*

**1. Introduction.** Our purpose here is to give a brief account of the results obtained over many years in the attempts to create a complex domain analogue to a very familiar theorem of real analysis. The theorem is one which is often introduced in our first calculus courses. It states that between any two real zeros of a differential real function $f$ lies at least one critical point of $f$ (zero of its first derivative $f'$).

This theorem first appeared in a book published in 1691 by the French mathematician Michel Rolle. Its publication had predated the adoption of the geometric representation of complex numbers by about 140 years. For, though this representation was devised by the Norwegian cartographer Caspar Wessel in 1797 and again by the Swiss mathematician Jean Argand in 1806, its universal acceptance had to await its invention in 1831 by none other than the great Karl Friedrich Gauss.

With this representation came the concepts of a complex variable $z = x + iy$ and a function of a complex variable. It is not surprising that some early attention was directed towards constructing in the complex domain counterparts to well-known theorems of real analysis such as Rolle's theorem. However, the generalization of Rolle's theorem to the complex plane is not obvious or trivial, as the following two examples show.

First, take the function $f(z) = e^{zi} - 1$ which has zeros at $z = 0$ and at $z = 2\pi$. If Rolle's theorem were valid, at least one critical point would be situated on the interval $0 < x < 2\pi$. But $f'(z) = ie^{zi}$ so that $f'$ has no zeros whatsoever.

Secondly, take the polynomial $f(z) = (z^2 - 1)(z - i\sqrt{3})$ which has zeros at the vertices $z = \pm 1$, $z = i\sqrt{3}$ of an isosceles triangle. If Rolle's theorem were valid, $f$ would have a critical point on each side of this triangle. But $f'(z) = 3[z - (i/\sqrt{3})]^2$ so that $f'$ has a single zero at $z = i\sqrt{3}$, a point interior to the triangle.

As the second example shows, the concept of a critical point lying between two real zeros (i.e.,

---

Morris Marden, born in 1905 at Boston, Massachusetts, was educated at Harvard University receiving an A.B. (highest honors in mathematics) in 1925, an A.M. in 1927 and a Ph.D. (under J. L. Walsh) in 1928. As a National Research Fellow from 1928 to 1930 he did postdoctoral work at Wisconsin under E. B. Van Vleck, at Princeton under Einar Hille, at Zurich under George Pólya and at Paris under Paul Montel. Later he studied applied mathematics at Brown under Stefan Bergman. Since 1930 he has been on the University of Wisconsin-Milwaukee faculty; since 1975 as UWM Distinguished Professor Emeritus. During 1975-77 he served as Visiting Distinguished Professor at California Polytechnic State University in San Luis Obispo. He has served as an MAA visiting lecturer and has lectured abroad extensively. He has two sons: Albert, professor of mathematics at the University of Minnesota; and Philip, a pediatrician practicing in Oconomowoc and Watertown, Wisconsin.

on the line segment joining the two zeros) generally is replaced in the complex plane by the concept of a critical point situated in some region containing the zeros of the given function. In our second example that region is a closed triangle but in our later examples a polygon or a circular disk may be the most convenient choice.

As to the importance of locating the critical points of a given function $f$, it is to be recalled that, whereas for a function of a real variable the determination of the critical points helps in locating the maxima and minima, for a function of a complex variable $f$ analytic in a region $T$ finding the critical points aids in determining where the map of $T$ by $w = f(z)$ fails to be conformal.

**2. Locating all the critical points.** An immediate corollary of Rolle's Theorem is the following result.

THEOREM 2.1. *If a real polynomial $f$ of the real variable $x$ has only real zeros all situated on the interval $I$: $a \leqslant x \leqslant b$ of the real axis, then all the critical points of $f$ also lie on $I$.*

It is Theorem 2.1 that was first to be generalized to the complex domain. The earliest such result was due to Gauss who in 1836 stated the following [**2**, pp. 21–24].

THEOREM 2.2 (Gauss). *The critical points of a polynomial $f$ which are not multiple zeros of $f$ are located at the equilibrium positions in a certain field of force. This field is one due to a particle placed at each zero of $f$, having a mass equal to the multiplicity of the zero and attracting according to the inverse distance law.*

From Gauss' Theorem we may deduce immediately the following fact.

THEOREM 2.3 (Lucas). *The critical points of a polynomial $f$ lie in the convex hull $H$ of the zeros of $f$.*

The convex hull $H$ of the zeros of $f$ is defined as the smallest convex polygon enclosing all the zeros of $f$.

Though Theorem 2.3 is a corollary of Theorem 2.2, nowhere in Gauss' papers does one find either a statement or proof of Theorem 2.3. The first published statement and proof of Theorem 2.3 appears to be due to the French engineer F. Lucas in 1874.

Another form of Lucas' Theorem is the following result.

THEOREM 2.4. *The critical points of a polynomial $f$ lie in any circle $C$ enclosing all the zeros of $f$.*

In fact, Theorem 2.4 is equivalent to Theorem 2.3. For, on the one hand, $C$ encloses $H$ and on the other hand, $H$ is the intersection of all circular disks covering $H$.

**3. Separation of the critical points.** If, instead of being given that all the zeros of polynomial $f$ lie in a single circle $C$, one presupposes that the zeros are distributed over a set of given circular disks, one may expect more specific results on the location of the critical points than given in Theorem 2.4. During the period 1918-1922 Joseph L. Walsh discovered a number of such results, two of which are the following [**2**, pp. 89-92].

THEOREM 3.1 (Walsh). *If a polynomial $f$ of degree $n$ has $m_1$ zeros in a circle $C_1$, with center at $z = c_1$ and radius $r_1$ and the remaining $m_2 = n - m_1$ zeros in a circle $C_2$ with center at $z = c_2$ and radius $r_2$, then every critical point of $f$ that does not lie in $C_1$ or $C_2$ lies in a third circle $C$ with center at $z = (m_2c_1 + m_1c_2)/n$ and radius $r = (m_2r_1 + m_1r_2)/n$.* (See Fig. 1.)

THEOREM 3.2 (Walsh). *If $C_0, C_1, \dots, C_p$ is a set of circles having a common external center 0 of similitude and if $f$ is a polynomial of degree $n$ that has $m_k$ zeros in circle $C_k$, $k = 0, 1, \dots, p$, then every critical point of $f$, not in a circle $C_k$, lies in a circle $C_k'$ where $C_1', C_2', \dots, C_q'$ is a set of circles also having 0 as an external center of similitude.* (See Fig. 2.)

An interesting new application of Theorem 3.1 is to a real polynomial $f$ with only real zeros.

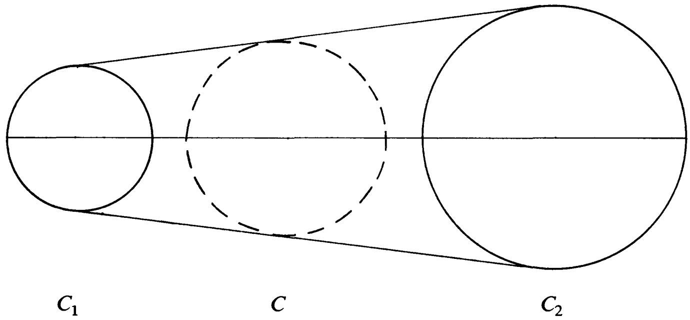



FIG. 1

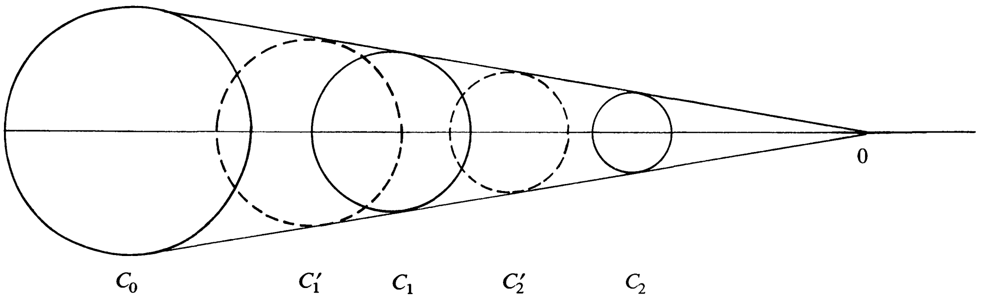



FIG. 2

THEOREM 3.3. *Let $f$ be a real polynomial of degree $n$ with only real zeros of which $m_1$ are located on the interval $I_1$: $a_1 \leqslant x \leqslant b_1$ and the remaining $m_2 = n - m_1$ are located on the interval $I_2$: $a_2 \leqslant x \leqslant b_2$, with $a_2 > b_1$. Then any critical point of $f$ not on $I_1$ or $I_2$ is located on the interval $I$: $a \leqslant x \leqslant b$, where $a = (m_2 a_1 + m_1 a_2)/n$ and $b = (m_2 b_1 + m_1 b_2)/n$.*

This theorem supplements Rolle's Theorem. To prove it, let us draw the circle $C_1$ whose center is at $c_1 + i\gamma$, where $c_1 = (a_1 + b_1)/2$, and which passes through the two points $z = a_1$ and $z = b_1$. Also let us draw the circle $C_2$ whose center is at $c_2 + i\gamma\lambda$, where

$$c_2 = (a_2 + b_2)/2 \quad \text{and} \quad \lambda = (b_2 - a_2)/(b_1 - a_1),$$

and which passes through the points $z = a_2$ and $z = b_2$. Then according to Theorem 3.1, any critical point of $f$ not in circle $C_1$ or circle $C_2$ lies in circle $C$ whose center is at

$$z = (1/n)[m_1 c_2 + m_2 c_1 + i\gamma(\lambda m_1 + m_2)]$$

and which cuts the real axis in the points $z = a$ and $z = b$. Now, as $\gamma$ is allowed to vary from $-\infty$ to $+\infty$, the intersection of all disks $C_1$ is interval $I_1$, the intersection of all disks $C_2$ is the interval $I_2$ and finally the intersection of all the circles $C$ is the interval $I$. Thus, we have proved Theorem 3.3.

**4. Further separation of the critical points.** Theorem 3.2 has been generalized to an arbitrary set of circular disks $C_k$, one that does not necessarily have a common external center of similitude [**2**, pp. 96-106].

THEOREM 4.1 (Marden). *For $j = 0, 1, 2, \ldots, p$ let $f_j(z)$ be a polynomial of degree $m_j$ that has all its zeros on the circular disk $C_j(z) \leqslant 0$, where $C_j(z) \equiv |z - c_j|^2 - r_j^2$. Then, if a critical point of the product*

$$f(z) = f_0(z) f_1(z) \cdots f_p(z) \tag{4.1}$$

*is not situated in one of the circular disks $C_0, C_1, \ldots, C_p$, it lies in a simply-connected region bounded*

*by an oval of the p-circular* $2p$*-ic curve* $E(z)=0$ *where*

$$\frac{E(z)}{\prod_0^p C_j(z)} = \sum_{j=0}^{p} \frac{nm_j}{C_j(z)} - \sum_{\substack{j=0\\k=j+1}}^{p} \frac{m_j m_k \tau_{jk}}{C_j(z)C_k(z)} \tag{4.2}$$

*and where*

$$\tau_{jk} = |c_j - c_k|^2 - (r_j - r_k)^2, \quad n = m_0 + m_1 + \cdots + m_p.$$

It is to be noted that, if $\tau_{jk} > 0$, then $\tau_{jk}$ equals the square of the length of the external common tangent of the circles $C_j(z) = 0$ and $C_k(z) = 0$.

To clarify Theorem 4.1, let us consider the special case $p = 2$, that is, of just three given circles $C_j(z) = 0$, $j = 0, 1, 2$. Equation (4.2) then simplifies to

$$E(z) \equiv n^2(x^2 + y^2)^2 + \text{lower degree terms}, \tag{4.3}$$

so that $E(z) = 0$ is the equation of a bicircular quartic. It may happen that (4.3) is factorable, thus:

$$E(z) = (x^2 + y^2 + \alpha_1 x + \beta_1 y + \gamma_1)(x^2 + y^2 + \alpha_2 x + \beta_2 y + \gamma_2), \tag{4.4}$$

in which case the bicircular quartic $E(z) = 0$ degenerates into a pair of circles as for Theorem 3.2.

As an example of a non-degenerate case, let us choose the circles $C_0$, $C_1$ and $C_2$ each of radius $r$ but with centers respectively $c_0 = 2i$, $c_1 = -3 - i$, $c_2 = +3 - i$, thus at the vertices of an isosceles triangle. Then (see Fig. 3) for $r = 0$, the bicircular quartic reduces to the points $z = \pm\sqrt{2}$. For $0 < r < \sqrt{3} - 1$, the quartic consists of two distinct ovals, one surrounding point $z = -\sqrt{2}$ and the other point $z = +\sqrt{2}$. For $r = \sqrt{3} - 1$ the two ovals touch one another at $z = 0$. For $\sqrt{3} - 1 < r < \sqrt{3} + 1$ the quartic is a single oval. For $r = \sqrt{3} + 1$, the quartic consists of a single oval and the isolated point $z = 0$. For $r > \sqrt{3} + 1$, the quartic consists of two ovals, with one enclosed in the other; in this case, any critical point, not in one of the given circular regions $C_0$, $C_1$ or $C_2$, lies in or on the outer oval of the bicircular quartic.

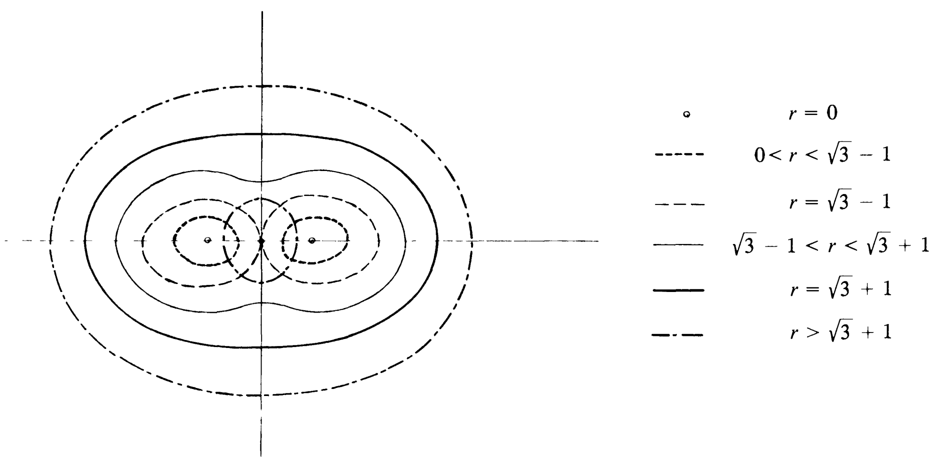



FIG. 3

In the general case equation (4.2) has the form

$$E(z) \equiv n^2(x^2 + y^2)^p + \text{lower degree terms}, \tag{4.5}$$

so that $E(z) = 0$ represents a $p$-circular $2p$-ic curve. While such curves have been known for

some time, very little has been written about their properties. It may happen that the curve $E(z)=0$ degenerates into a set of circles as in the case of Theorem 3.2.

Another example where curve $E(z)=0$ degenerates into a set of circles is that in which the circles $C_j$ are all of the same radius $r$ and have their centers at the $p+1$ vertices of a regular polygon $P$. The curve $E(z)=0$ then reduces to a set of concentric circles centered at the center of $P$.

**5. Locating only some critical points.** So far we have discussed some complex domain analogues, not to Rolle's Theorem, but to its corollary, Theorem 2.1, that involves all the zeros and all the critical points of a polynomial. However, Rolle's Theorem proper deals with the location of the critical points of a real polynomial in relation only to pairs of real zeros.

A complex domain analogue closer to Rolle's Theorem is the following due to the two British mathematicians J. H. Grace in 1902 and P. J. Heawood in 1907 [**2**, pp. 107-110].

THEOREM 5.1. *If $z_1$ and $z_2$ are any two distinct zeros of a polynomial $f$ of degree $n$, then at least one critical point of $f$ lies on the circular disk $|z-c|\leqslant r$, where*

$$c=(z_1+z_2)/2 \quad \textit{and} \quad r=(|z_1-z_2|/2)\cot(\pi/n).$$

This result turns out to be the best possible in that the limits given in Theorem 5.1 are attained by the polynomial

$$f(z)=\int_{-1}^{z}[s-i\cot(\pi/n)]^{n-1}\,ds.$$

From Theorem 5.1 one may deduce the following result, discovered independently by the American topologist Alexander in 1915.

THEOREM 5.2. *If the disk $|z|\leqslant R$ contains two zeros of an nth degree polynomial $f$, the concentric disk $|z|\leqslant R\csc(\pi/n)$ contains at least one critical point of $f$.*

In an attempt to generalize Theorem (5.2), the Japanese mathematician Kakeya established in 1917 the following result [see **2**, pp. 113-121].

THEOREM 5.3. *If the disk $|z|\leqslant R$ contains $p$ zeros of an nth degree polynomial where $2\leqslant p\leqslant n$, then there exists a number $\phi(n,p)$ dependent only upon $n$ and $p$ such that at least $p-1$ critical points of $f$ lie in the disk $|z|\leqslant R\phi(n,p)$.*

The value $\phi(n,n)=1$ follows from Lucas' result in Theorem 2.4. Kakeya himself was able to show $\phi(n,2)=\csc(\pi/n)$ confirming our Theorem 5.2, but admitted an inability to calculate $\phi(n,p)$ for general $p$. In 1927 the Polish mathematician M. Biernacki calculated the value $\phi(n,n-1)=[1+(1/n)]^{1/2}$. In 1936 Marden showed that

$$\phi(n,p)\leqslant\csc(\pi/2q), \tag{5.1}$$

where $q=n-p+1$, and in (1945) Biernacki showed that

$$\phi(n,p)\leqslant\prod_{k=1}^{n-p}[(n+k)/(n-k)]. \tag{5.2}$$

However, neither (5.1) nor (5.2) provides the least upper bound to $\phi(n,p)$.

In 1972, Marden [**4**] studied the Kakeya problem relative to the special family of polynomials

$$f(z)=(z-z_1)^{\lambda}(z-z_2)^{\mu}(z-z_3)^{\nu},$$

where

$$\lambda+\mu=p,\quad \nu=n-p,\quad |z_1|=|z_2|=R,\quad \text{and}\quad |z_3|\geqslant R.$$

For such polynomials he proved that $\phi(n,p)\leqslant[2-(p/n)]^{1/2}$ and that this bound is attained when $p$ is an even integer.

An application of the above leads to the result:

THEOREM 5.4 (Marden). *An nth degree polynomial P is at most p-valent* ($2 \leqslant p \leqslant n$) *on the disk* $D$: $|z| \leqslant \sin[\pi/2(n-p)]$ *if, in the unit disk* $|z| \leqslant 1$, *P has at most* $p-1$ *critical points.*

For a function $f$ to be at most $p$-valent in a region $S$ means that, if $A$ is an arbitrary real or complex number, the equality $f(z) = A$ holds on at most $p$ distinct points $z$ within $S$.

Theorem 5.4 follows almost immediately from Theorem 5.3 and inequality (5.1). For, if for some number $A$ the equation $P(z) = A$ were valid at $p+1$ points in the disk $D$, then $f(z) = P(z) - A$ would have at least $p$ critical points in the disk

$$|z| \leqslant \phi(n, p+1)\sin[\pi/2(n-p)] \leqslant \csc[\pi/2(n-p)]\sin[\pi/2(n-p)] = 1,$$

in contradiction to the hypothesis of Theorem 5.4. Hence, $P(z) = A$ in at most $p$ points of $D$.

**6. Some recent conjectures.** During the past dozen years considerable interest has been aroused regarding the location of the critical points of any polynomial $f$ all of whose zeros lie in the unit disk $|z| \leqslant 1$. By Theorem 2.4 we know that all the critical points of $f$ also lie on the unit disk. The question recently raised is how close to each zero do the critical points lie [see **3**].

The following conjecture was made by the Bulgarian mathematician B. Sendov in 1962 but became later known as the "Ilyeff Conjecture."

CONJECTURE 1. *If all the zeros of an nth degree polynomial lie on the disk* $|z| \leqslant 1$ *and* $z_0$ *is any one of the zeros, then at least one critical point of f lies within unit distance from* $z_0$.

This conjecture has been proved in a number of special cases including $n = 2, 3, 4, 5$, but not as yet in general.

Another conjecture was made in 1969 by the American mathematicians Goodman and Roth, the Canadian-Indian mathematician Rahman and the German mathematician Schmeisser.

CONJECTURE 2. *Under the same hypotheses as for Conjecture* 1, *at least one critical point of f lies on the disk* $|z - (z_0/2)| \leqslant 1 - (|z_0|/2)$.

This has been proved when $|z_0| = 1$, but some counterexamples have been devised for the case $|z_0| < 1$ by M. J. Miller [see **5**]. One of these is the sixth degree polynomial $P(z) = (z - 0.84)Q(z)$, where

$$Q(z) = z^5 + 1.182303196z^4 + 1.34007004z^3 + 1.34007004z^2 + 1.182303196z + 1.$$

Miller shows that the six zeros of $P(z)$ are approximately the following quantities:

$$z_0 = 0.84, \quad z_1 = -1, \quad z_2 = 0.415548 + 0.909571i,$$
$$z_3 = -0.506700 + 0.862122i, \quad z_4 = \bar{z}_2, \quad z_5 = \bar{z}_3.$$

Obviously $|z_0| < 1$. His calculations show that

$$|z_1| = |z_2| = |z_3| = |z_4| = |z_5| = 1.$$

On the other hand, the five zeros of $P'(z)$ are computed by him to be approximately the following:

$$\zeta_1 = -0.162505, \quad \zeta_2 = -0.159819 + 0.055911i, \quad \zeta_3 = 0.098445 + 0.485713i,$$
$$\zeta_4 = \bar{\zeta}_2, \quad \zeta_5 = \bar{\zeta}_3.$$

If Conjecture 2 were valid for polynomial $P$ and for the above zero $z_0$, then the inequality

$$S_k = |\zeta_k - 0.42| \leqslant 0.58$$

would hold for at least one $k$, $k = 1, 2, \ldots, 5$. However, his calculations show that $S_k \geqslant 0.582505$ for all $k$, thus refuting Conjecture 2.

**7. Critical points of non-polynomials.** So far we have surveyed the results aimed at extending Rolle's Theorem to polynomials in a complex variable. We now shall consider the corresponding work with entire functions, rational functions and more general functions of a complex variable.

The names of Laguerre, Cesàro, and Pólya are associated with the earlier work on entire functions of a complex variable. By an entire function is, of course, meant one for which the Maclaurin series

$$f(z) = b_0 + b_1 z + b_2 z^2 + \cdots$$

converges for all values of $z$. Of special interest are entire functions of finite order $\rho$, meaning

$$\rho = \limsup_{k \to \infty} \frac{k \log k}{\log\,(1/|b_k|)} = \limsup_{r \to \infty} \frac{\log \log M(r,f)}{\log r} < \infty,$$

where $M(r,f) = \{\max |f(z)|,\ |z| = r\}$.

The following is a result about the critical points of entire functions of finite order $\rho$.

THEOREM 7.1 (Marden). *If $f$, an entire function of finite order $\rho$, has all its zeros in a semi-infinite strip $T$, at most $n$ critical points of $f$ lie in sector $S$ (see* Fig. 4), *where $n$ is the largest positive integer less than or equal to $\rho$.*

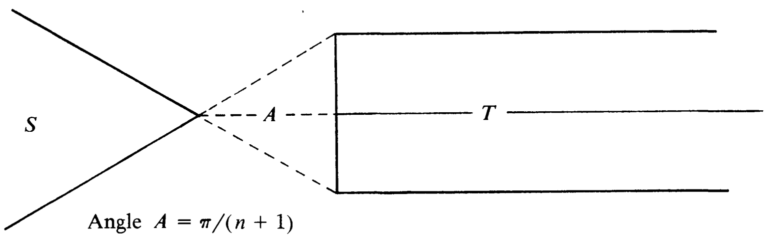



FIG. 4

For example, let $\rho$ be a non-negative integer and let

$$f(z) = e^{\lambda z^{\rho}} \prod_{k=0}^{m} \left[(z-k)^2 + 1\right].$$

This entire function of order $\rho$ has all its zeros in a semi-infinite strip $T$: $0 \leqslant x$, $-1 \leqslant y \leqslant 1$. Hence, at most $n = \rho$ critical points lie in the sector $S$.

Another example

$$f(z) = \prod_{k=-p}^{k=p} \cos(z - ki)^{1/2}$$

is that of an entire function of order $\rho = 1/2$. Here one finds as the semi infinite strip $T$, $x \geqslant \pi^2/4, |y| \leqslant p$, whereas $n = 0$. Thus no critical points lie in $S$. In fact, one can show that all the critical points lie in $T$.

More generally, it may be proved that in the case of any entire function $f$ with an order $\rho$, $0 \leqslant \rho < 1$, any convex polygon $H$ which contains all the zeros of $f$ also contains all the critical points of $f$, just as in Theorem 2.3. However, $H$ will ordinarily be unbounded, since $f$ may have an infinite number of zeros.

For rational and meromorphic functions, there are theorems similar to those on polynomials and entire functions respectively.

For a more general function $f$ which is analytic in a region $T$, some results have been obtained by the noted French mathematician Jean Dieudonné in 1930 [see **1**]. Supposing that $f$ has in $T$ a zero at $z = z_1$ and one at $z = z_2$, and that the circle $C$ with the line segment from $z_1$ to $z_2$ as diameter is contained in $T$, he then determines conditions on $f$ such that at least one critical point of $f$ lies in $C$. However, these conditions are too complicated to be stated here.*

*If $z_1$ and $z_2$ are sufficiently close, there should be a critical point in their immediate vicinity, since the critical points are continuous functions of the zeros. For, $z_2 \to z_1$ makes $z_1$ a double zero and thus a critical point.

**8. Concluding remarks.** In summary, we have surveyed the efforts made towards devising some complex plane counterparts to Rolle's Theorem. These attempts began about one hundred and fifty years ago with Gauss and have continued to the present day. None of the results so far have had the generality and simplicity of Rolle's Theorem. Hence, it remains a challenge for the future to find a true analogue of Rolle's Theorem in the complex domain.

**Acknowledgments.** The author wishes to thank UWM Teaching Assistant Sherrie Nicol for the drawings accompanying this paper, and his colleague Professor David H. Schultz for the numerical calculations.

## References

1. Jean Dieudonné, Sur une généralisation du théorème de Rolle aux fonctions d'une variable complexe, Ann. of Math., 31 (1930) 79–116.
2. Morris Marden, Geometry of Polynomials, Math. Survey no. 3, Amer. Math. Soc., Providence, 1966.
3. ______, Conjectures on the critical points of a polynomial, this MONTHLY, 90 (1983) 267–276.
4. ______, Kakeya's problem for polynomials, Tensor, N.S. 25 (1972) 365–371.
5. Michael J. Miller, On Ilieff's Conjecture and the Geometry of Polynomials, Doctoral Thesis, University of Rochester, 1984.

# ON ALL SORTS OF AUTOMORPHISMS*

PETER HILTON
*Department of Mathematical Sciences, SUNY at Binghamton, Binghamton, NY 13901*

**0. Introduction.** If $\phi$ is a linear transformation of a finite-dimensional vector space into itself, then there is a simple criterion that $\phi$ be invertible, that is, an automorphism, given by $\det\phi \neq 0$. However, this criterion fails, as we shall see, in general for an endomorphism of a finitely-generated module over a commutative unitary ring. In this note we find a criterion for such an endomorphism $\phi$ to be an automorphism in terms of the existence of certain polynomial annihilators of $\phi$, basing ourselves on a Cayley-Hamilton Theorem for such endomorphisms. We then examine the very different situation which occurs if we drop the condition that the module be finitely-generated; and this in turn leads us to criteria for special classes of automorphisms of (non-commutative) groups.

The criterion in terms of polynomial annihilators, and the extension to the non-finitely-generated case, arose, in the context of homotopy theory, in trying to understand some interesting work of J. M. Cohen [2], [3] on self-maps of fiber spaces. Cohen, in fact, used one aspect of the criterion in [2], and gave an ad hoc proof, in the case of abelian groups, in [3]. However, the problem of understanding Cohen's work was exacerbated by a defect in his definition of a *pseudo-identity* of an abelian group as a special type of automorphism. Nevertheless, it became plain, from a study of [3], that a simplification of a basic tool used by Cohen could be achieved by using a

---

Peter Hilton was born in London, England, on April 7, 1923. He was educated at Oxford University and spent the years 1942–5 of the Second World War as a cryptanalyst at Bletchley Park, where he became very friendly with J. H. C. (Henry) Whitehead. On demobilization he returned to Oxford to obtain a D.Phil. in algebraic topology under Whitehead's supervision. After an academic career in England, culminating in his holding the Mason Chair of Pure Mathematics at the University of Birmingham, he came to the U.S. in 1962 to become Professor of Mathematics at Cornell University. He is now Distinguished Professor of Mathematics at SUNY-Binghamton.

Hilton's research interests are, as they have always been, in algebraic topology and homological algebra, but he has taken an increasing interest in mathematics education, at all levels, since coming to the U.S. He has been Chairman of the U.S. Commission on Mathematical Instruction and Secretary of the International Commission. He was First Vice-President of the MAA, 1978–80.

*This paper is based on the text of the 17th Harry M. Gehman Lecture, delivered at the meeting of the Seaway Section of the Mathematical Association of America, Broome Community College, April 7, 1984.

Cayley-Hamilton Theorem for endomorphisms of finitely-generated abelian groups or, more generally, finitely-generated modules over a commutative unitary ring.

We include in Section 1 a proof of such a generalized Cayley-Hamilton Theorem for completeness; it is, of course, not new (see, for example, [7]). However, it seems worth drawing attention to the fact that such a proof is not available if one confines attention to vector spaces, since the operator domain $\Lambda$ must, in any case, be embedded in the polynomial ring $\Lambda[t]$. Such a process is fundamental to the proofs given in [**1**], [**8**], [**11**].

We use the Cayley-Hamilton Theorem to study polynomial annihilators of an endomorphism $\phi$ of a (finitely-generated) module $A$ over a commutative ring $\Lambda$. We enumerate 4 assertions about $\phi$, each of which implies that $\phi$ is an automorphism and discuss their mutual implications. If $A$ is finitely generated, they are all equivalent to $\phi$ being an automorphism—this follows from the Cayley-Hamilton Theorem—but if $A$ is not finitely generated, then no two are equivalent and all are stronger than the assertion that $\phi$ is an automorphism. These last results, obtained with the use of readily accessible examples, appear in Section 2. In fact, Cohen's notion of pseudo-identity is equivalent to assertion (d) in Theorem 1.2.

In the course of our discussion we find characterizations equivalent to assertions (b) and (d) which immediately generalize to the non-abelian case. Thus we are ready to pass, not merely from vector spaces to modules, but even to groups with operators. This direction is precisely that taken by Roitberg and the author in [**5**], [**6**], [**10**].

We should, perhaps, apologize for perpetrating the term 'demonic polynomial' in this paper. The term 'monic polynomial,' is standard (though we here permit the slight generalization of allowing the leading coefficient to be $\pm 1$); the term 'comonic polynomial' for a polynomial with constant term $\pm 1$ seems reasonable in view of the dual role it plays in our theory. To describe a polynomial which is both monic and comonic as 'bimonic' would be an etymological solecism. To stop short at 'dimonic' would be perverse and cowardly!

In reference [**9**] a nice proof is given that an endomorphism of a finitely-generated module over a commutative ring is annihilated by a monic polynomial, but the proof does not allow us to minimize its degree (see Remark (ii) following Theorem 1.2).

**1. The Cayley-Hamilton Theorem.** Let $\Lambda$ be a unitary commutative ring and let $A$ be a finitely-generated (left) $\Lambda$-module, with finite generating set $\mathbf{a} = \{a_1, a_2, \ldots, a_n\}$. Given any endomorphism $\phi\colon A \to A$, we may associate with $\phi$ the matrix

$$M = \begin{pmatrix} \lambda_{11} - t & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} - t & \dots & \lambda_{2n} \\ \vdots & & & \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nn} - t \end{pmatrix}$$

over the polynomial ring $\Lambda[t]$, where

$$\phi a_i = \sum_{j=1}^{n} \lambda_{ij} a_j, \quad i = 1, 2, \ldots, n, \quad \lambda_{ij} \in \Lambda. \tag{1.1}$$

Then $\det M$, the determinant of the matrix $M$, is an element of $\Lambda[t]$, called the *characteristic polynomial of $\phi$ relative to the equations* (1.1). Let us write $F(t)$ for this polynomial. We may then prove (see, for example, [**7**]):

THEOREM 1.1 (Generalized Cayley-Hamilton Theorem). $F(\phi)A = 0$.

REMARK. If $A$ is a free module and $\Lambda$ is a ring such that a free finitely generated $\Lambda$-module has a unique rank, then the characteristic polynomial of $\phi$ is uniquely determined by $\phi$. This remark, of course, covers the case of a finite-dimensional vector space.

We include a proof of Theorem 1.1 for completeness; it will be just an adaptation of the standard proof proceeding via Cramer's rule for solving systems of linear equations (see [**7**]).

*Proof of Theorem.* Set $R = \Lambda[t]$ and make $A$ into an $R$-module by defining $ta = \phi a, a \in A$. Then from (1.1) we infer that $\Sigma_{j=1}^{n} c_{ij} a_j = 0, i = 1, 2, \ldots, n$, where

$$c_{ij} = \begin{cases} \lambda_{ij}, & i \neq j, \\ \lambda_{ii} - t, & i = j. \end{cases}$$

Regarding these as equations in the quantities $a_j$, we infer that $F(t)a_k = 0$, $k = 1, 2, \cdots, n$. But since $\mathbf{a}$ is a generating set for $A$, this means that $F(t)A = 0$. Finally $F(\phi)A = 0$ since $ta = \phi a$.

We note that $F(t)$ is a polynomial of degree $n$ whose leading coefficient is $\pm 1$; we call such a polynomial *monic*. We also introduce the term *comonic* for a polynomial whose constant term is $\pm 1$, and *demonic* for a polynomial which is both monic and comonic. Note that the set of monic (comonic, demonic) polynomials is closed under multiplication. We may now use Theorem 1.1 to prove the following result.

THEOREM 1.2. *Let $\Lambda$ be a unitary commutative ring and let $A$ be a $\Lambda$-module. Consider the following assertions about an endomorphism $\phi$: $A \to A$:*

(a) *$\phi$ is an automorphism;*
(b) *$\forall a \in A$, $\exists$ comonic $F(t)$ such that $F(\phi)a = 0$;*
(c) *$\exists$ comonic $F(t)$ such that, $\forall a \in A$, $F(\phi)a = 0$;*
(d) *$\forall a \in A$, $\exists$ demonic $F(t)$ such that $F(\phi)a = 0$;*
(e) *$\exists$ demonic $F(t)$ such that, $\forall a \in A$, $F(\phi)a = 0$.*

*Then*

$$\begin{array}{ccccc} & & \text{(d)} & & \\ & \nearrow & & \searrow & \\ \text{(e)} & & & & \text{(b)} \to \text{(a)}. \\ & \searrow & & \nearrow & \\ & & \text{(c)} & & \end{array} \tag{1.2}$$

*Moreover, if $A$ is finitely generated, all 5 assertions are equivalent.*

*Proof.* All implications in (1.2) are trivial, except for (b) $\to$ (a). Now, given (b), we have an equation, for each $a \in A$,

$$\left(c_0\phi^n + c_1\phi^{n-1} + \cdots + c_{n-1}\phi \pm 1\right)a = 0.$$

This equation shows immediately that $a \in \phi A$ and that $a = 0$ if $\phi a = 0$. Thus $\phi$ is an automorphism.

Now assume $A$ finitely generated. If $\mathbf{a} = \{a_1, a_2, \ldots, a_n\}$ is a generating set, then there is a characteristic polynomial of $\phi$, say $G(t)$, which is monic of degree $n$, such that $G(\phi)A = 0$. If $\phi$ is an automorphism then there is also a characteristic polynomial of $\phi^{-1}$, say $\overline{G}(t)$, also monic of degree $n$, such that $\overline{G}(\phi^{-1})A = 0$. Notice that $t^n\overline{G}(t^{-1})$ is comonic of degree $\leqslant n$. Then if we set

$$F(t) = tG(t) + t^n\overline{G}(t^{-1}),$$

it is easy to see that $F(t)$ is demonic (of degree $n + 1$) and $F(\phi)A = 0$, so that (e) holds. In the light of (1.2), this proves the equivalence of all 5 assertions if $A$ is finitely generated.

REMARKS (i). The equivalence of (a) and (e), in the finitely generated case, was the starting point of Cohen's investigations in **[2]**, **[3]**.

(ii) We noted above that the polynomial $f(t)$ we constructed had degree $(n + 1)$. Thus the construction produces a polynomial of degree $(n + 1)$, where $n$ is the minimum number of elements required to generate $A$. It is clear that we cannot, in general, do better than this. For example, let $A = \mathbb{Z}/5$, generated by $a$, and let $\phi$: $A \to A$ be given by $\phi a = 2a$. There is plainly no demonic linear polynomial $F(t)$ such that $F(\phi)a = 0$. However the quadratic polynomial $t^2 + 1$ will evidently serve.

(iii) We were careful to state, in our argument, that the degree of $t^n\overline{G}(t^{-1})$ is less than or equal to $n$. It would be less than $n$ if the constant term of $\overline{G}(t)$ were zero. Now the constant term of the characteristic polynomial of $\phi$ relative to the equations (1.1) is $\det(\lambda_{ij})$, so the question arises whether we can have $\det(\lambda_{ij}) = 0$ when $\phi$ is an automorphism. The following example is, perhaps, a little surprising.

Let $A$ be the abelian group $\mathbb{Z}/2 \oplus \mathbb{Z}/3 = (a, b)$ and define $\phi: A \to A$ by

$$\phi a = 3a - 3b, \qquad \phi b = 2a - 2b. \tag{1.3}$$

Then $\phi$ is the identity but the matrix associated with (1.3) is $\begin{pmatrix} 3 & -3 \\ 2 & -2 \end{pmatrix}$, whose determinant is zero. Notice that, in this example, our generating set is minimal in the weak sense that no element may be omitted, but not in the strong sense that its cardinality is minimal.

We have been able to prove, in this direction, the following very partial result. In the statement and proof we use the term 'minimal generating set' in the *strong* sense.

THEOREM 1.3. *Let $A$ be a finitely-generated module over the principal ideal domain $\Lambda$, let $\{a_1, a_2, \dots, a_n\}$ be a minimal generating set for $A$, and let $\phi$: $A \to A$ be given by*

$$\phi a_i = \sum_{j=1}^{n} \lambda_{ij} a_j, \quad i = 1, 2, \dots, n, \quad \lambda_{ij} \in \Lambda. \tag{1.4}$$

*Then if $\phi$ is an automorphism*, $\det(\lambda_{ij}) \neq 0$.

*Proof*. Let us first choose a special minimal generating set $(a_1, a_2, \dots, a_k, b_1, b_2, \dots b_l)$, $k + l = n$, for $A$, where $a_i$ is of order $\mu_i$, $i = 1, 2, \dots, k$, with $\mu_1 | \mu_2 | \cdots | \mu_k, \mu_1 > 1$, and $\{b_1, b_2, \dots, b_l\}$ is a basis for a free $\Lambda$-module. Thus $\mu_1, \mu_2, \dots, \mu_k$ are the *invariant factors* for $A$. Let $T$ be the torsion submodule of $A$. We then have a commutative diagram

$$\begin{array}{ccc} T \rightarrowtail & A \twoheadrightarrow & A/T \\ \downarrow \phi' & \downarrow \phi & \downarrow \phi'' \\ T \rightarrowtail & A \twoheadrightarrow & A/T \end{array}$$

where $\phi', \phi, \phi''$ are automorphisms. If we choose $\{a_1, a_2, \dots, a_k\}$ as generating set for $T$ and $\{b_1 T, b_2 T, \dots, b_l T\}$ as generating set for $A/T$, then $\det\phi = (\det\phi')(\det\phi'')$, where $\det\phi$ has its obvious meaning. Certainly $\det\phi'' \neq 0$, so it remains to show that $\det\phi' \neq 0$. Thus, for this part of the argument, we may assume that $A$ is a torsion module, and hence that $k = n$. Now if $\phi$: $A \to A$ is the identity, then, in (1.4),

$$\lambda_{ij} \equiv 0 \bmod \mu_j, \quad i \neq j; \quad \lambda_{ii} \equiv 1 \bmod \mu_i.$$

It follows that $\det(\lambda_{ij}) \equiv 1 \bmod \mu_1$, so that $\det(\lambda_{ij}) \neq 0$.

Now let $\phi$ be an arbitrary automorphism of the torsion module $A$ with special minimal generating set $\{a_1, a_2, \dots, a_n\}$. Then we have, say,

$$\phi a_i = \sum \lambda_{ij} a_j, \qquad \phi^{-1} a_i = \sum \overline{\lambda}_{ij} a_j,$$

and so $(\det\lambda_{ij})(\det\overline{\lambda}_{ij}) \neq 0$. Thus $\det\lambda_{ij} \neq 0$, and this proves our assertion if we had chosen a special minimal generating set.

We now turn to the general case—and, of course, no longer assume that $A$ is a torsion module. Let (1.4) be given, and suppose we have a special minimal generating set $(\bar{a}_1, \bar{a}_2, \dots, \bar{a}_n)$. Then

$$\bar{a}_i = \sum p_{ij} a_j, \qquad a_i = \sum q_{ij} \bar{a}_j,$$

so that

$$\phi \bar{a}_i = \sum \overline{\lambda}_{ij} \bar{a}_j,$$

where $(\overline{\lambda}_{ij}) = (p_{ij})(\lambda_{ij})(q_{ij})$. Thus

$$\det(p_{ij})\det(\lambda_{ij})\det(q_{ij}) = \det(\overline{\lambda}_{ij}) \neq 0,$$

so that $\det(\lambda_{ij}) \neq 0$ and the theorem is proved.

It is probably not easy to improve this result; Craig Squier has produced an example of a Noetherian integral domain $\Lambda$, a $\Lambda$-module $A$ generated by 2 elements, and an automorphism $\phi$: $A \to A$ admitting a zero determinant. It should be noted, however, that, in this example, $\Lambda$ is not a unique factorization domain.

Squier's example* is easy to describe, but it is not easy to establish its crucial property! Thus, let $a, b, c, d$ be indeterminates over the complex numbers $\mathbb{C}$ and let $\Lambda$ be obtained from $\mathbb{C}[a, b, c, d]/(ad - bc)$ by adjoining an inverse of $a + d$. Let $M$ be the $\Lambda$-module generated by $(u, v)$ with the relations $cu = av$, $du = bv$; and, finally, let $\phi$: $M \to M$ be multiplication by $a + d$. Then we may write $\phi u = au + bv$, $\phi v = cu + dv$, so that the associated matrix is $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with zero determinant. The crucial property to establish is that $M$ is not a cyclic module. This requires either a long and involved computational proof or an appeal to the methods of algebraic geometry. Indeed, it turns out that $\Lambda$ is the coordinate ring of a 3-dimensional quasi-affine algebraic variety over $\mathbb{C}$, and hence a Noetherian integral domain.

**2. Non-finitely generated modules.** We revert to the 5 assertions of Theorem 1.2 and show, by examples, that they are all distinct if $A$ is no longer assumed finitely generated. To validate our examples we need the following propositions.

Let $\phi: A \to A$ be an endomorphism of the $\Lambda$-module $A$. For each $a \in A$ let $H_a = H_a(\phi)$ be the submodule generated by the orbit of $a$ under $\phi$, thus

$$H_a = H_a(\phi) = \langle a, \phi a, \dots, \phi^n a, \dots \rangle.$$

PROPOSITION 2.1. *Let $A$ be a $\Lambda$-module and $\phi$: $A \to A$. Then the following assertions are equivalent*:

(i) *Assertion* (b) *of Theorem* 1.2;
(ii) $\forall a \in A, a \in \phi H_a$;
(iii) $\forall B \subseteq A$ *such that* $\phi B \subseteq B, \phi B = B$.

*Proof*. The equivalence of (i) and (ii) is almost immediate. Now if (iii) holds, then, since obviously $\phi H_a \subseteq H_a$, it follows that $\phi H_a = H_a$, so $a \in \phi H_a$. Conversely, given (ii), let $a \in B$. Then $\phi a \in B$ and so, inductively, $\phi^n a \in B$ for all $n$. Thus $H_a \subseteq B$ and $a \in \phi H_a \subseteq \phi B$, so that $\phi B = B$.

PROPOSITION 2.2. *Let $A$ be a $\Lambda$-module and let $\phi$: $A \to A$ be an automorphism. Then*

$$a \in \phi^{-1} H_a(\phi^{-1}) \leftrightarrow H_a(\phi) \text{ is finitely generated.}$$

*Proof*. If $H_a(\phi)$ is finitely generated then $\exists n$ such that $\phi^n a \in \langle a, \phi a, \dots, \phi^{n-1} a \rangle$. But then

$$a \in \langle \phi^{-n} a, \phi^{-(n-1)} a, \dots, \phi^{-1} a \rangle \subseteq \phi^{-1} H_a(\phi^{-1}).$$

Conversely if $a \in \phi^{-1} H_a(\phi^{-1})$, then

$$a \in \langle \phi^{-n} a, \phi^{-(n-1)} a, \dots, \phi^{-1} a \rangle,$$

for some $n$, so that $\phi^n a \in \langle a, \phi a, \dots, \phi^{n-1} a \rangle$. But then an easy induction shows that

$$\phi^m a \in \langle a, \phi a, \dots, \phi^{n-1} a \rangle,$$

for all positive $m$, so that

$$H_a(\phi) = \langle a, \phi a, \dots, \phi^{n-1} a \rangle.$$

PROPOSITION 2.3. *Let $A$ be a $\Lambda$-module and $\phi$: $A \to A$. Then assertion* (d) *of Theorem* 1.2 *is equivalent to the joint assertion that, $\forall a \in A, a \in \phi H_a$ and $H_a$ is finitely generated.*

*The author is grateful to Craig Squier for permission to publish this preview of a forthcoming article.

*Proof.* Let $F(t)$ be a demonic polynomial such that $F(\phi)a = 0$. Then, say,

$$\left(\phi^n + c_1\phi^{n-1} + \cdots + c_{n-1}\phi \pm 1\right)a = 0.$$

This makes plain that $a \in \phi H_a$ and that $\phi^n a \in \langle a, \phi a, \ldots, \phi^{n-1}a\rangle$. As above, this last implies that $H_a$ is finitely generated. Conversely suppose that for all $a \in A$, $a \in \phi H_a$ and $H_a$ is finitely generated. Since $a \in \phi H_a$ it follows from Proposition 2.1 that $\exists$ comonic $F(t)$ such that $F(\phi)a = 0$. Since $H_a$ is finitely generated, it follows from Propositions 2.1 and 2.2 that $\exists$ comonic $\bar{F}(t)$ such that $\bar{F}(\phi^{-1})a = 0$. Choose $n$ sufficiently large that $t^n\bar{F}(t^{-1})$ is a monic polynomial in $t$ with zero constant term and degree greater than that of $F(t)$. Then $G(t) = F(t) + t^n\bar{F}(t^{-1})$ is a demonic polynomial such that $G(\phi)a = 0$.

We are now ready to prove by examples that the 5 assertions of Theorem 1.2 are all distinct if $A$ is not assumed finitely generated. In all our examples $A$ will be an abelian group.

EXAMPLE 2.1. Consider the automorphism $\mathbb{Q} \xrightarrow{\frac{1}{2}} \mathbb{Q}$. Clearly (c) holds with $F(t) = 2t - 1$. However (d) is false since $H_a$ is not finitely generated unless $a = 0$.

EXAMPLE 2.2. Let $\phi_n\colon \mathbb{Z}^n \to \mathbb{Z}^n$ be an automorphism with minimum polynomial of degree $n$ (for example, the cyclic permutation $(a_1, a_2, \ldots, a_n) \to (a_2, \ldots, a_n, a_1)$ of generators); and let $\phi = \oplus\phi_n\colon \bigoplus_n \mathbb{Z}^n \to \bigoplus_n \mathbb{Z}^n$. Then (d) is true. For any element of $\bigoplus_n \mathbb{Z}^n$ belongs to a finitely generated subgroup on which $\phi$ is an automorphism, so Theorem 1.2 itself ensures the truth of (d). On the other hand (c) is false since no nonzero polynomial in $\phi$ can vanish on every element of $\bigoplus_n \mathbb{Z}^n$.

EXAMPLE 2.3. Let $\phi_i\colon \mathbb{Q} \to \mathbb{Q}$ be multiplication by $\frac{1}{i}$ and let $\phi = \oplus_i\phi_i\colon \oplus_i\mathbb{Q} \to \oplus_i\mathbb{Q}$. Then (b) is true, since any element of $\oplus_i\mathbb{Q}$ belongs to some subgroup $\oplus_{i=1}^{j}\mathbb{Q}$ and is thus annihilated by the comonic polynomial $\prod_{i=1}^{j}(it - 1)$. On the other hand (d) is false since it was even false in the simpler Example 2.1. Moreover, (c) is also false. For if there existed a comonic polynomial $F(t)$ such that $F(\phi)A = 0$, then there would exist a monic polynomial $G(t)$ such that $G(\phi^{-1})A = 0$. But if $\alpha_i$ is the unity of the $i$th copy of $\mathbb{Q}$ then $\phi^{-1}(\alpha_i) = i\alpha_i$, so we would have $G(i) = 0$. This would have to be true for every positive integer $i$, an evident contradiction.

EXAMPLE 2.4. Consider the automorphism $\mathbb{Q} \xrightarrow{2} \mathbb{Q}$. It is evident from Proposition 2.1 that (b) is false since $\phi\mathbb{Z} \subseteq \mathbb{Z}$ but $\phi\mathbb{Z} \neq \mathbb{Z}$.

It is probably sensible, when $A$ is not finitely generated, to concentrate on assertions (a), (b) and (d) of Theorem 1.2. It is then surely worth observing that the equivalence of (ii) and (iii) in Proposition 2.1, as also the validity of Proposition 2.2, do not depend on $A$ being commutative. Thus we may consider the equivalent formulations of (b) and (d), in terms of $H_a(\phi)$, in studying endomorphisms of groups or endomorphisms of $\pi$-groups. (Here we understand by a *$\pi$-group* a group $G$ on which a fixed group $\pi$ acts, all homomorphisms of $\pi$-groups to respect the $\pi$-action). This is the point of view to be found, for example, in [5], [**6**], [**10**]. In fact, Proposition 2.3 establishes that assertion (d) is the most significant of the 5 assertions of Theorem 1.2 in the non-finite-generated case, since the equivalent assertion about $H_a$ leads to the important consequences in the generalized context.

**3. Appendix.** In [4] the author proves a generalized Cayley-Hamilton Theorem in the following form.

THEOREM. *Over an arbitrary commutative ring, if matrices $A$ and $B$ commute, then $\chi_A(B) \equiv 0 \bmod B - A$, where $\chi_A(t)$ is the characteristic polynomial of $A$.*

We observe that, if $\phi, \psi\colon M \to M$ are commuting endomorphisms of a $\Lambda$-module $M$, then, for *any* polynomial $F(t)$, $F(\phi) - F(\psi) \equiv 0 \bmod \phi - \psi$. Thus for any polynomial $F(t)$ such that

$F(\phi)M = 0$, it follows that

$$F(\psi)M \equiv 0 \bmod (\phi - \psi)M.$$

The theorem above is thus an immediate consequence of this observation and Theorem 1.1.

**References**


1. M. Braun, Differential Equations and Their Applications, 3rd Ed., Springer-Verlag, 1983.
2. J. M. Cohen, A spectral sequence automorphism theorem; applications to fibre spaces and stable homotopy, Topology, 7 (1968) 173–177.
3. ______, Clarification to a result, Topology, 9 (1970) 299–300.
4. Marvin Jay Greenberg, Note on the Cayley-Hamilton Theorem, this MONTHLY, 91 (1984) 193–195.
5. Peter Hilton, A note on finitary automorphisms, Aspects of Mathematics and its Applications, North Holland, 1985 (to appear).
6. Peter Hilton and Joseph Roitberg, On pseudo-identities I, Archiv der Mathematik, 41 (1983) 204–214.
7. Kenneth Hoffman and Roy Kunze, Linear Algebra, 2nd ed., Prentice-Hall.
8. Saunders Mac Lane and Garrett Birkhoff, Algebra, 2nd Ed., Macmillan, 1979.
9. Maurice Orzech, Onto endomorphisms are isomorphisms, this MONTHLY, 78 (1971) 357–362.
10. Joseph Roitberg, Finitary automorphisms and integral homology, Topological Topics, London Mathematical Society Lecture Notes, 86 (1983) 164–168.
11. Gilbert Strang, Linear Algebra and its Applications, 2nd Ed., Academic Press, 1980.


---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.
*For instructions about submitting Notes for publication in this department see the inside front cover.*

## GENERALIZED LEAST SQUARES AND EIGENVALUES

ALAN SHUCHAT
*Department of Mathematics, Wellesley College, Wellesley, MA 02181*

In a recent note in this section of the MONTHLY [**2**], Hung C. Li discusses the generalized least squares problem of finding the best-fitting line $L$ to a set of points in the plane. Here "generalized" means that the sum $S$ of squares of the *perpendicular* distances from the points to $L$ should be minimized, as opposed to ordinary least squares where distances are measured vertically or horizontally. This forms the basis of orthogonal regression [**3**] and principal component analysis [**4**]. Li's approach is to treat $S$ as a function of the two parameters that determine $L$ and to minimize it by partial differentiation. However, we can obtain additional geometric insight and extend the solution more easily to higher dimensions by using linear algebra, as in [**4**]. The presentation given here is simpler and accessible to undergraduates.

If $L$ is a line and $x$ a point in $R^n$, we let $d(x, L)$ denote the orthogonal distance from $x$ to $L$. Given points $x^1, \dots, x^m$ in $R^n$, we seek a line $L$ for which the sum $S = \Sigma_k d(x^k, L)^2$ is minimal. We will show that $L$ must pass through the centroid of the given points and that its direction can be given by any eigenvector for the largest eigenvalue of a certain symmetric matrix. More generally, the $d$ largest eigenvalues determine a best-fitting affine subspace of dimension $d$.

We first show that the desired line must pass through the centroid, which for convenience we assume to be at the origin, i.e., $\Sigma_k x^k = 0$. Let $L'$ be any line $x = w + tv$ in $R^n$, where $v$ is a unit vector, and let $L$ be its translation $x = tv$ to the origin. For each $x$ in $R^n$,

$$d(x, L)^2 = \|x\|^2 - (x \cdot v)^2,$$

**Telegraphic Reviews**

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**Precalculus, T*(13: 1).** College Algebra, Third Edition. Karl J. Smith, Patrick J. Boyle. Brooks/Cole, 1984, x + 502 pp. [ISBN: 0-534-03747-X] Contains various applications in different fields such as psychology and physics. Each chapter introduced with an historical note--a chapter overview. Chapter material introduced primarily through examples. A few non-traditional problems are included along with the many standard textbook exercises. (TR, Second Edition, January 1983.) TR

**Precalculus, T(13: 1).** Trigonometry, Third Edition. Margaret L. Lial, Charles D. Miller. Scott Foresman, 1985, xiv + 351 pp, $22.95. [ISBN: 0-673-18016-6] Straightforward, basic trigonometry text. Contains a few interesting application problems. Some explanation included with each topic. Approach is mainly through examples. (First Edition, TR, November 1977; Second Edition, TR, August-September 1981.) TR

**Precalculus, T*(13: 1).** Technical Mathematics. Rudolph E. Lynn. Wiley, 1985, xvi + 712 pp, $31.95. [ISBN: 0-471-88743-9] A good pre-calculus text for engineering/technical students. Contains many specific technical applications. Has good explanations; very readable; emphasis is on application of rules. TR

**Precalculus, T(13: 1).** College Algebra, Second Edition. Bernard Kolman, Arnold Shapiro. Academic Pr, 1985, xiii + 492 pp, $21. [ISBN: 0-12-417897-9] Excellent up-date of a fine text (First Edition, TR, February 1982). Covers equations, functions, exponential and logarithmic functions, conics, matrices, sequences. Optimal section on linear programming. New: anecdotal digressions (very nice), and answers to selected problems. MA

**Precalculus, T(13).** Trigonometry. Roland E. Larson, Robert P. Hostetler. DC Heath, 1985, xvi + 378 pp, $24.95. [ISBN: 0-669-08611-8] Chapter material presented primarily through written explanations and augmented by proofs and/or examples. Nice treatment of sine/cosine graphing. Does not need the attention devoted to calculators, particularly the separate calculator exercises--these should be integrated. Contains additional topics such as polar graphing. TR

**Precalculus, T(13: 1).** College Algebra, Ninth Edition. Paul K. Rees, Fred W. Sparks, Charles Sparks Rees. McGraw-Hill, 1985, xiii + 545 pp, $28.95. [ISBN: 0-07-051735-5] Suitable for a pre-calculus course. Has a wide variety of applications. Does not include material on circular functions. Emphasis is on examples and some problem solving techniques. (Sixth Edition, TR, June-July 1972; Seventh Edition, TR, August-September 1977; Eighth Edition, TR, October 1981.) TR

**Precalculus, T*(13: 1).** Precalculus Mathematics, Second Edition. Thomas W. Hungerford, Richard Mercer. Saunders College, 1985, xx + 663 pp. [ISBN: 0-03-000843-1] Presentation of standard precalculus topics is clear and very readable. Changes in this edition include new material on conic sections, systems of equations, and sequences and sums, together with new review material and many new exercises. (First Edition, TR, March 1981.) JRG

**Foundations, T(18: 1), P.** Constructibility. Keith J. Devlin. Perspectives in Math. Logic. Springer-Verlag, 1984, xi + 425 pp, $58. [ISBN: 0-387-13258-9] Constructible set theory is an extension of Zermelo-Fraenkel set theory in which the notion of what constitutes set is precisely defined. This reference text presents a comprehensive account of theory of constructible sets. Covers Gödel hierarchy in first part and Jensen hierarchy in second. Short exercise sets. KS

**Foundations, T(13-14), S, L.** Gödel's Theorem Simplified. Harry J. Gensler. U Pr of America, 1984, iii + 83 pp, $6.75 (P); $12.75. [ISBN: 0-8191-3869-X; 0-8191-3686-1] Non-technical exposition of proof of Gödel's first incompleteness theorem: arithmetic cannot be reduced to any axiomatic system. First shows that lower arithmetic, consisting of propositions without quantifiers, can be axiomatized. Explains logical terms and symbolism. Some comments on philosophical implications of theorem. KS

**Graph Theory, T(16-17: 1, 2).** Graphs, Groups and Surfaces. Arthur T. White. Math. Stud., V. 8. Elsevier Sci, 1984, xiii + 314 pp, $30 (P). [ISBN: 0-444-87643-X] The nine chapters of the First Edition (TR, August-September 1974; Extended Review, January 1976) have been revised and updated; six new chapters have been added covering voltage graphs, non-orientable imbeddings, hypergraph imbeddings, map automorphism groups, and change ringing (fascinating application). JRG

**Topological Groups, T(18: 1, 2), S, P.** Ergodic Theory and Semisimple Groups. Robert J. Zimmer. Mono. in Math., V. 81. Birkhauser Boston, 1984, x + 209 pp, $29.95. [ISBN: 0-8176-3184-4] Very readable account, in a research-expository style, of recent results of G.A. Margulis and of the author, on amenable group actions, rigidity, arithmetic lattices, Lie groups, and a careful introduction to ergodic theory, beginning with the definition of ergodic actions. Includes numerous examples and an extensive bibliography, but no exercises. YN

**Calculus, T(14: 1, 2).** Intermediate Calculus, Second Edition. Murray H. Protter, Charles B. Morrey, Jr. Undergrad. Texts in Math. Springer-Verlag, 1985, x + 648 pp, $38. [ISBN: 0-387-96058-9] Thorough coverage of multivariable calculus topics from the geometry of $R^3$ through vector integration. Includes chapters on infinite series (including Fourier series) and appended material on matrices and determinants. Flexible treatment allows for a variety of syllabi. There are few sophomore-level texts available--this is a welcome re-entry. JRG

**Calculus, T*(15: 1, 2), L.** Advanced Calculus, Third Edition. Wilfred Kaplan. Addison-Wesley, 1984, xiv + 721 pp, $36.95. [ISBN: 0-201-11680-4] More proofs and exercises than in the Second Edition (TR, February 1974); some rewriting; new material on potential theory. Presupposes calculus, provides introductory linear algebra. Rigorous as before; much emphasis on physical motivation and applications. Many references to the literature. DFA

**Calculus, T*(13: 2), S.** Calcul différentiel et intégral 1: Fonctions réelles d'une variable réelle. Jacques Douchet, Bruno Zwahlen. Pr Polytechniques Romandes, 1983, ix + 244 pp, Sfr. 68 (P). [ISBN: 2-88074-026-6] Appropriate for freshman honors' calculus courses (in French). Provides concise and simple yet rigorous proofs of all the standard theorems (Heine-Borel-Lebesgue, Bolzano-Weierstrass, Rolle, Dini). Also includes some special topics, for instance Minkowski's inequality. Postpones all multivariate calculus to the second volume. Many examples and numerous exercises, some leading to clever proofs of classical facts. Thin, yet substantial: good preparation for real analysis. YN

**Calculus, T(13: 3), S.** Vorlesungen zur höheren Mathematik. Johannes Weissinger. Bibliographisches Institut, 1984. Band 1: Grundlagen, Differentialrechnung, Hochschultaschenbücher, B. 613, 190 pp [ISBN: 3-411-00613-7]; Band 2: Integralrechnung, Lineare Algebra, Vektorrechnung, Hochschultaschenbücher, Band 614, 260 pp, (P). [ISBN: 3-411-00614-5] Course intended for engineers (in German). Carefully explains the standard topics in differential and integral calculus, with discrete and numerical methods when appropriate, e.g., Newton's method in two variables. Presents many constructive or computational proofs, e.g., the binomial theorem, and states without proof the harder results, e.g., the convergence of Fourier series. Excellent as lecture notes, perhaps terse as text. YN

**Real Analysis, T(16-17: 1, 2), S, P, L.** Undergraduate Analysis. Serge Lang. Undergrad. Texts in Math. Springer-Verlag, 1983, xiii + 545 pp, $28. [ISBN: 0-387-90800-5] Aimed at upper-level undergraduates, the book begins with a brisk survey of elementary calculus using techniques "not emphasized in elementary courses." Part two is a discussion of uniform convergence in several contexts including the Stone-Weierstrass theorem and the Riemann integral. Part three concerns applications including Fourier series and integrals. Part four is differential calculus in n-space and part five is multiple integration concluding with an introduction to differential forms. Some exercises, mostly extending the theory or on special topics. Appendix on algebra. Index. JS

**Real Analysis, P.** Lecture Notes in Mathematics-1067: Absolute Summability of Fourier Series and Orthogonal Series. Yasuo Okuyama. Springer-Verlag, 1984, vi + 118 pp, $7.50 (P). [ISBN: 0-387-13355-0] Focuses on absolute Nörlund summability and absolute Riesz summability of Fourier series and orthogonal series. Dry, with awkward grammer in places. BH

**Complex Analysis, P.** American Mathematical Society Translations, Series 2, Volume 122: Ten Papers on Complex Analysis. M.A. Lavrent'ev, et al. AMS, 1984, vi + 222 pp, $54. [ISBN: 0-8218-3081-3] Translations of Russian research papers on a variety of topics in complex analysis. BH

**Numerical Analysis, S(16-17), P*.** A New Approach to Scientific Computation. Ed: Ulrich W. Kulisch, Willard L. Miranker. Notes & Reports in Comp. Sci. & Appl. Math., No. 7. Academic Pr, 1983, xv + 384 pp, $39. [ISBN: 0-12-428660-7] 12 papers, mostly by members of the Institute for Applied Mathematics at Karlsruhe, West Germany, on the development and use of a computer arithmetic for intervals of reals, complex numbers, vectors and arrays. For many operations and numerical problems, optimal accuracy can be obtained. Implementation in languages such as Pascal-SC. RWN

**Numerical Analysis, T*(16: 2), S, L*.** Computational Methods for Partial Differential Equations. E.H. Twizell. Ser. in Math. & Its Applic. Halsted Pr, 1984, 276 pp, $57.95. [ISBN: 0-470-27511-1] Explains the classical methods of finite differences and finite elements, and compares their relative merits. Applies them to the trilogy of elliptic, parabolic, and hyperbolic partial differential equations. Discusses characteristics, consistency, and convergence; includes theoretical and computational exercises. Instructors may add complements in linear algebra, existence and uniqueness, mesh generation, and physical motivations (largely omitted topics). YN

**Functional Analysis, P.** *Modèles étalés des espaces de Banach*. B. Beauzamy, J.-T. Lapresté. Hermann, 1984, 210 pp, 160F (P). [ISBN: 2-7056-5965-X]

**Functional Analysis, P.** *Jordan Operator Algebras*. Harald Hanche-Olsen, Erling Størmer. Mono. & Stud. in Math., No. 21. Pitman, 1984, viii + 183 pp, $55. [ISBN: 0-273-08619-7] Assumes basic knowledge of Banach and Hilbert spaces. Develops algebraic theory of Jordan algebras and then studies JB algebras their weakly-closed analogous JBW algebras. (A JB algebra A is a Jordan Banach algebra satisfying $\|a^2\| = \|a\|^2$ and $\|a^2\| \leq \|a^2 + b^2\|$ for a, b in A.) Large bibliography. BH

**Analysis, P.** *Coloquio de Sistemas Dinamicos*. Ed: J.A. Seade, G. Sienra. Aportaciones Matematicas, V. 1. Sociedad Matematica Mexicana, 1985, viii + 167 pp, (P). First volume in a series of lecture notes sponsored by the Sociedad Matemática Mexicana. This volume features papers on dynamical systems, mostly in English, from a December 1983 conference at the Centro de Investigacíon en Matemáticas (CIMAT) in Guanajuato, Mexico. LAS

**Differential Geometry, S(18), P.** *Déformations Infinitésimales des Structures Conformes Plates*. Jacques Gasqui, Hubert Goldschmidt. Progress in Math., V. 52. Birkhauser Boston, 1984, 226 pp, $24.95. [ISBN: 0-8176-3260-3] Research monograph presenting the authors' recent results. Constructs and studies resolutions of sheaves of Killing fields over conformally flat, real, Riemannian manifolds of any dimension, using tools from the theory of overdetermined systems of partial differential equations. YN

**Differential Geometry, T?(17: 2), S*.** *Differential Manifolds*. Serge Lang. Springer-Verlag, 1985, ix + 230 pp, $19.80 (P). [ISBN: 0-387-96113-5] Introduces sprays and vector bundles over infinite dimensional manifolds. Also proves such classical results as the Morse-Palais lemma and Stokes' theorem with singularities. Yet students might fail to appreciate the depth of such statements as "Hironaka tells me that by using the resolution of singularities [...]" (p. 204). No exercises. Certainly original, but perhaps not an introductory text. (1972 Addison-Wesley hardcover edition, TR, August-September 1972.) YN

**Probability, T(16-18: 3), S, P, L.** *Probability*. A.N. Shiryayev. Transl: R.P. Boas. Springer-Verlag, 1984, xi + 577 pp, $48. [ISBN: 0-387-90898-6] This textbook is based on a three-semester course given by the author at Moscow State University. The first part is devoted to the elementary theory; the second part to the modern theory ($\sigma$-algebras, measures, Lebesgue integral, characteristic functions); the third part to random processes with discrete parameters (Markov chains, theory of martingales). Thorough treatment with a large number of examples and exercises. LCL

**Probability, P*.** *Classical Potential Theory and Its Probabilistic Counterpart*. J.L. Doob. Grund. der Math., V. 262. Springer-Verlag, 1984, xxv + 846 pp, $58. [ISBN: 0-387-90881-1] "The purpose of this book is to develop (the) correspondence between potential theory and probability theory by examining in detail classical potential theory, that is, the potential theory of Laplace's equation, together with the corresponding probability theory, that is Martingale theory." No knowledge of potential theory is presupposed. Reader should be familiar with basic probability concepts through conditional expectations. The necessary lattice theory, set theory and capacity theory are covered in Appendices. JK

**Statistics, T(17: 1), S.** *The Modern Forecaster: The Forecasting Process Through Data Analysis*. Hans Levenbach, James P. Cleary. Lifetime Learning, 1984, xiv + 537 pp, $31.50. [ISBN: 0-534-03361-X] The authors present forecasting as a process rather than as a set of disjoint techniques. There are five parts to the book: an introduction to the forecasting process, exploratory data analysis techniques, smoothing and regression techniques, the econometric approach to forecasting, and time series modeling using Box-Jenkins methodology. MT

**Statistics, S?(14-17).** *Experimentation and Statistical Validation*. Norbert L. Enrick. Robert Krieger, 1983, 121 pp, $14.50. [ISBN: 0-89874-445-8] Based on a series of articles which appeared in the *Journal of Neurological and Orthopaedic Surgery*. Describes various statistical procedures, using illustrations from the health and related sciences, in an attempt to bring an understanding and appreciation of the contributions that can be made by statistics. RSK

**Statistics, T(17: 1), P*.** *Analysis of Survival Data*. D.R. Cox, D. Oakes. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1984, ix + 201 pp, $25. [ISBN: 0-412-24490-X] This book gives a thorough introduction to the analysis of survival data, also referred to as failure-time data. It can be used either as an introductory text for the subject or as a reference for applied statisticians. Includes discussion of parametric and nonparametric analysis techniques for both univariate data sets and data sets with explanatory variables. MT

**Statistics, P.** *Nonparametric Functional Estimation*. B.L.S. Prakasa Rao. Prob. & Math. Stat. Academic Pr, 1983, xiv + 522 pp, $70. [ISBN: 0-12-564020-X] Attempts to bring together literature about nonparametric estimation of density functions, distribution functions, regression functions, failure rates, etc. Most of the results have not appeared in a book before; there are bibliographic notes at the end of each chapter. MT

**Computer Literacy, T(13-14), S.** *Computer Information: A Modular System*. Marjorie M. Leeson. SRA, 1985, xvi + 591 pp, (P). [ISBN: 0-574-21720-7] A very complete, flexible computer literacy text. Contains interesting commentary on issues and trends as well as a brief introduction to BASIC for selected micros. Each of four modules can be used independently from others--a nice feature. TR

**Computer Programming, T(15), S.** BASIC, A Modular Approach, Second Edition. Robert G. Thompson. Charles E. Merrill, 1985, xi + 366 pp, (P). [ISBN: 0-675-20280-9] Programming in BASIC with good style. Concepts presented in a clear, organized way; examples good. Two possible problems: programming is presented too early (no algorithm design) and extended BASIC has not been used, with resultant lower code readability. FA

**Computer Programming, T(15: 1), S.** FORTRAN 77 for Scientists and Engineers, Second Edition. J.N.P. Hume, R.C. Holt. Reston, 1985, xvi + 364 pp, (P). [ISBN: 0-8359-2065-8] Structured programming with examples mainly from science and engineering. Includes design of algorithms using "structured English" and Fortran-77. Builds up to some fairly complex algorithms and data structures. Appendix presents an introduction to UNIX and VAX/VMS. FA

**Computer Programming, T(13-18: 1).** BASIC Programming for the Apple Computer. Robert M. Bateson, Robin D. Raygor, Gregory Bitz. West, 1985, xx + 278 pp, $14.95 (P) [ISBN: 0-314-85290-5]; Instructor's Manual to Accompany, x + 199 pp, (P). [ISBN: 0-314-87247-7] A good beginning text with a well-written instructor's manual. Contains many examples and problems for each topic. Includes comments on possible errors and how to correct them. TR

**Computer Programming, T(13-14: 1), S.** Applied Structured BASIC. Roy Ageloff, Richard Mojena. Ser. in Computer Information Systems. Wadsworth, 1985, xvi + 550 pp, (P). [ISBN: 0-534-04740-8] Standard introductory text for BASIC. Does contain a section on modules and matrix operations. Allows for system variations. Contains applications geared to many fields. TR

**Applications (Actuarial Science), P.** Loss Distributions. Robert V. Hogg, Stuart A. Klugman. Wiley, 1984, x + 235 pp, $29.95. [ISBN: 0-471-87929-0] There are five chapters in this book: the first chapter is an introduction; the second and third chapters discuss probabilistic models and methods of fitting these models to data; and the fourth and fifth chapters discuss choosing and applying distributions of the size of insurance losses. MT

**Applications (Biology), P.** Lecture Notes in Biomathematics-56: Gonorrhea, Transmission Dynamics, and Control. Herbert W. Hethcote, James A. Yorke. Springer-Verlag, 1984, ix + 105 pp, $9 (P). [ISBN: 0-387-13870-6] Presentation of mathematical models for the incidence and transmission of gonorrhea in several populations, and evaluation of disease control procedures. JRG

**Applications (Communication Theory), P.** Applications of Walsh and Related Functions: With an Introduction to Sequency Theory. K.G. Beauchamp. Academic Pr, 1984, xvi + 308 pp, $55. [ISBN: 0-12-084180-0] Update on developments in signal theory in the past decade. Background chapters in sequence theory are followed by applications to signal processing in one and two dimensions, to communications, and in the logical description of digital designs. Extensive list of references accompanies each chapter. JRG

**Applications (Economics), P.** Lecture Notes in Economics and Mathematical Systems-236: A Disequilibrium Model of Real and Financial Accumulation in an Open Economy. Giancarlo Gandolfo, Pietro Carlo Padoan. Springer-Verlag, 1984, vi + 172 pp, $11.50 (P). [ISBN: 0-387-13889-7]

**Applications (Engineering), S(17), P.** Lecture Notes in Mathematics-1086: Sensitivity of Functionals with Applications to Engineering Sciences. Ed: V. Komkov. Springer-Verlag, 1984, v + 130 pp, $7.50 (P). [ISBN: 0-387-13871-4] The functionals in question weakly formulate boundary value problems, e.g., buckling beams, and involve design parameters such as shape (as opposed to control parameters such as loads). Design modifications, e.g., boundary shapes altered by vector fields, change the functionals and weak solutions. Finding stationary functionals leads to optimal designs, e.g., strongest cross-sectional shapes; mathematically, the unknown is the boundary value problem itself, rather than its solution. YN

**Applications (Information Theory), T**(16-17: 1), S*, P*, L.** Mathematics of Kalman-Bucy Filtering. P.A. Ruymgaart, T.T. Soong. Ser. in Inform. Sci., V. 14. Springer-Verlag, 1985, x + 170 pp, $29. [ISBN: 0-387-13508-1] This superb, mathematically sound and accessible exposition covers stochastic differential equations and processes, with mean-square calculus in Hilbert spaces. Begins with essentials of functional analysis, measure, and integration theory. Ends with Kalman-Bucy filters: recursions for reconstructing estimators of random vectors observed with random noise. Gives exercises and answers. Outstanding text, both for engineers and for "pure" mathematicians. YN

## Reviewers

MA: Melissa Anderson, St. Olaf; FA: Fahrad Anklesaria, Macalester; DA: David Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; YN: Yves Nievergelt, St. Olaf; AO: Arnold Ostebee, St. Olaf; TR: Teresa Reardon, St. Olaf; AWR: A. Wayne Roberts, Macalester; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

## Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

### Seaway Section

The spring meeting of the Seaway Section held April 26-27, 1985 was a joint meeting with the New York State Mathematics Association of Two Year Colleges. Approximately 250 attended the meetings.

Invited Addresses:

"Is Discrete Mathematics the New Math of the 80's?" by Anthony Ralston, SUNY Center at Buffalo.
"A Vector Approach to Euclidean Geometry," (The Gehman Lecture) by Wolfgang Jurkat, Syracuse University.

Short Presentations:

"Alternatives to Least Square Estimators," by Marvin Gruber, Rochester Institute of Technology.
"Left, Right, and Other Things," by Saleh I. Assad, Erie Community College.
"Derivations and Commutativity," by Howard Bell, Brock University.
"Models Based on the Mean and the Median," by David Farnsworth, Rochester Institute of Technology.
"Mathematics for Computer Science--A First Course," by Eric R. Muller, Brock University.

Student Papers:

"Function Spaces," by Jay Robartes, State University College at Potsdam.
"There's More Than One Way to Slice a Space," by Dwight Tuinstra, State University College at Potsdam.
"Extensions of the Heisenberg-Weyl Inequality," by Martin Smith, McMaster University.
"A Predictive Model for the Relationship Between Age and Memory Span," by Wendy Witt, Rochester Institute of Technology.
"Sabermetrics," by James Johnston, St. Bonaventure University.
"Analysis of Antithetic Variates Technique Applied to Cosage," by Thomas Johnson, Rochester Institute of Technology.

### Michigan Section

The annual spring meeting of the Michigan Section was held on May 3-4, 1985 at Western Michigan University in Kalamazoo, Michigan. Meeting attendance was approximately 140.

Invited Lectures:

"Another Approach to Riemann-Stieltjes Integrals," by Ken Ross, University of Oregon.
"Some Recent Problems and Results in Combinatorial Geometry," by Paul Erdos, Hungarian Academy of Sciences.
"Mathematics Programs in Russia," by Richard Dahlke, University of Michigan.
"Statistical Analysis of the Love Canal Environmental Study," by Michael Stoline, Western Michigan University.
"Editing the Monthly," by Mary and Robert Wardrop, Central Michigan University.
"Bernoulli's Polynomials and Periodic Splines," by Dennis Pence, Western Michigan University.
"The Recent Solution of the Bieberbach Conjecture," by Peter Duren, University of Michigan.
"On the Napoleon Triangle Theorem," by John Wetzel, University of Illinois.
"Recent Results in Linear Programming," by Michael Saks, Rutgers University and Bell Communication Research.

Short Presentations:

"Discrete Mathematics," by Karen Sharp, Mott Community College.
"The Joint ACM/MAA Retraining Program in Computer Science," by Philip Demavois, Northwestern Michigan College.
"From Computer Science to Topology," by Linda Brinn, University of Michigan, Dearborn.
"Curriculum: A Need for Change," by Terry Wesner, Henry Ford Community College.
"A Survey of K-12 Curriculum," by Harry Nustad, Henry Ford Community College.
"New Directions in Two-year Mathematics," by Keith Shuert, Oakland Community College; Karen Sharp, Mott Community College.
"On the Distribution of First Digits of Linear Sequences," by Peter Shive, Saginaw Valley State College.
"Toward a Michigan Mathematics Early Placement Test: A Progress Report," by John Kiltinen, Northern Michigan University.

Student Papers:

"Karmarkar's New Polynomial-time Algorithm for Linear Programming," by Jack Snoeyink, Calvin College.
"The Logistic Distribution as a Model to Describe the Joint Action of Drugs," by T.J. Widmar, Western Michigan University.
"Fractals and Computer Graphics," by Eric Jones, Michigan Technological University.
"Tarsky Geometry and New Theorem Proving Techniques," by Jon Mills, Western Michigan University.
"Development and Utilization of Logic Machine Control Language," by Jennifer Arbanas and Rich Stevens, Western Michigan University.

Panel Discussion:

International Congress of Mathematics Education.

## New Jersey Section

The spring meeting of the New Jersey Section was held as a joint meeting with the Mathematics Association of Two-Year Colleges of New Jersey on April 27, 1985 at Monmouth College, West Long Branch, New Jersey. Approximately 50 people attended.

Invited Addresses:

"Graph Theory: An Application of Path-length Distributions," by Edward Boyne, Montclair State.
"Mathematical Research and Artificial Intelligence," by Susan Epstein, Hunter College.
"Some Applications of Pure Mathematics at the Census Bureau," by Mary-Mulry Liggan, U.S. Census Bureau.
"Highlights of the 1985 International Congress of Mathematics Educators in Australia," by Gloria Gilmer, Copin State College.

## Northeastern Section

The spring meeting of the Northeastern Section was held on June 14-15, 1985 at Norwich College in Northfield, Vermont. Approximately 80 participants attended the meeting.

Invited Addresses:

"Why Every Manager Should Be a Mathematician," by Alfred Willcox, Executive Director of MAA, Washington, D.C.
"The Role of Mathematics and Problem Solving in Business Decision Making," by Gordon Prichett, Babson College.
"Public-Key Cryptosystems," by Dennis Luciano, Western New England College.
"A Discrete/Continuous Mathematics Course for Non-math Non-computer Science Majors," by Margaret Cozzens, Northeastern University.

Contributed Paper:

"Purely Geometric Analog of Lorentz-Einstein Equations Associated with the Unit Circle," by Herbert Nichol, Drexel University.

Student Papers:

"Discrete Prolate Spheroidal Sequences and Their Applications to Secure Voice Communications," by Gwyn J. Crouch, Worcester Polytechnic Institute.
"Estimating Age-specific Fecundity for Modeling Population Dynamics of Clams," by Ann Kochendorfer, Fairfield University.
"Estimating Optimal Transformations for Regression," by Skip Olson, Bowdoin College.
"Periodicity Within Pascal's Triangle Modulo k," by David A. Sawtelle, Colby College.

Workshop Presentations:

"Spreadsheets for Approximation Methods in Numerical Analysis," by Steve Wiitala, Norwich University.
"The International Conference on Mathematical Education of 1984," by Joan Mundy, University of New Hampshire.
"Management Information Systems," by Cynthia Lane, Bedford Research.

Panel Discussion:

"Undergraduate Mathematical Training and Its Use in an Industrial Setting," by Leonard J. Borucki (Moderator), IBM; Frederick Brown, Union Mutual Insurance Company; Joseph P. Noonan, Bedford Research; Karl Seer of Pratt-Whitney Aircraft.

## Texas Section

The annual meeting of the Texas Section was held at Texas A&M University at Galveston, Texas on April 11-13, 1985. There were 209 registrants for the meeting.

### Invited Addresses:

"Lacunary Fourier Series, from Sidon to Pisier," by Kenneth Ross, University of Oregon.
"The Examinations of the MAA Committee on American Mathematics Competitions," by Walter Mientka, University of Nebraska.
"Approximation of Matrices," by Paul R. Halmos, University of Santa Clara.

### Contributed Papers:

"A Theorem on Planar Continua and an Application to Automorphisms of the Field of Complex Numbers," by Robert R. Kallman, North Texas State University.
"Banach Algebras Satisfying $Ax^2 = Ax$," by Victor A. Belfi, Texas Christian University.
"Evaluating Special Limits Using Riemann Sums," by Sudhir K. Goel and Dennis Rodriquez, University of Houston, Downtown.
"New Characterizations Using Feebly Open and Regular Open Sets," by Charles Dorsett, Louisiana Tech University.
"Integration Techniques Using Parametric Integrals," by Aurel J. Zajta and Sudhir K. Goel, University of Houston, Downtown.
"The Cauchy-Bunyakovsky-Swartz Equality--1821 to the Present," by Robert S. Doran, Texas Christian University.
"An Interesting Subset of $\ell^{\infty}$," by Elizabeth M. Bator, North Texas State University.
"An Alternate Proof of the Mean Mean Value Theorem," by Vincent P. Schielack, Jr., Texas A&M University.
"The Problem of Non-T--Special Spaces," by Joseph C. Warndof, Midwestern State University.
"Kaplansky's Theorem on C(X) Revisited," by Don E. Edmondson, University of Texas at Austin.
"On Infinite Periodic Rings," by Effraim P. Armendariz, University of Texas at Austin.
"The Norm-closed Convex Hull of the Orbit of an Automorphism Group of a W*-algebra," by Joseph M. Szucs, Texas A&M University, Galveston.
"SDP Transforms of the Quartic Diophantine Equation of Euler," by Aurel J. Zajta, University of Houston, Downtown.
"On Equivalent Endomorphisms of Finite Groups," by John R. Durbin, University of Texas at Austin.
"Independence Spaces," by Wynne Johnson, University of Texas, Arlington.
"Generalized Absolute Neighborhood Retracts," by Stuart Anderson, East Texas State University.
"Painting a Fence Derives the Fundamental Theorem of Calculus," by Laurence Maher, North Texas State University.
"A Spiral Approach to Discrete and Continuous Mathematics Instruction," by Richard A. Alo and Michael G. Murphy, University of Houston, Downtown.
"Accreditation of Undergraduate Computer Science Programs," by Ashok Kumar and Sudhir K. Goel, University of Houston, Downtown.
"A New Mathematics Degree--Bachelor of Science in Quantitative Methods," by Ron Barnes, University of Houston, Downtown.
"In Search of Excellence--A Comparative Study in the Mathematics Teacher Program Between the Province of Taiwan, Republic of China, and the United States," by Ping-Tung Chang, Laredo State University.
"Technical Mathematics Program at Eastfield College," by Vivian Dennis, Eastfield College.
"Calculus Teaching and the Role of Testing," by Gregory D. Foley, Austin Community College.
"A Plea for Nothing," by Tammie Ann Hill, Prairie View A&M University.
"Projectionally Exposed Cones in Convex Programming," by George D. Poole, Lamar University.
"Uses of the Microcomputer for Mathematics Classroom Demonstrations," by Timothy P. Donovan, East Texas State University.
"An Interactive Graphics Package for Spline Approximation," by Susan Hunt, Baylor University.
"Third Order Iteration Techniques," by Donald F. Bailey, Trinity University.
"On the Oscillation of Solutions of Certain Third Order Difference Equations," by Beverly Smith, Texas Southern University.
"Solving Polynomial and Rational Function Inequalities," by Bella Wiener, Pan American University.
"The Microcomputer as an Aid in Teaching Business Math," by Carol Ghomi and Dennis Rodriquez, University of Houston, Downtown.
"Searching for Narcissistic Numbers by Computer," by John F. Lamb, Jr., East Texas State University.
"A Classroom Note on Numerical Singularity," by John Dennis, Rice University.

### Special Sessions:

"Forum on Funding Higher Education in Texas: Problems, Proposed Solutions, and Portents of the Future."

### Special Reports:

"Progress Report on New Mathematics Curriculum and New Standards for Mathematics Teacher Certification," by Barbara Montalto, Texas Education Agency.

Panel Discussion:

"Mathematics Course Work--Who's Calling the Shots?" by Rose Marie Smith (Moderator), Texas Women's University; John Allen, North Texas State University; Don Bailey, Trinity University; Don Edmondson, University of Texas, Austin; Peter Lindstrom, Northlake College.

## Wisconsin Section

The fifty-third annual meeting of the Wisconsin section was held at Marquette University in Milwaukee, Wisconsin, April 26-27, 1984. Approximately 175 people attended.

Invited Addresses:

"Some Unexpected Results in Elementary Mathematics," by Ivan Niven, University of Oregon.
"Peculiarities of Rational and Irrational Numbers," by Ivan Niven, University of Oregon.

Short Presentations:

"Algorithm for Computer Graphics," by Andrew Matchett, University of Wisconsin, LaCrosse.
"How Does One Treat an Ill-Posed Problem," by Robert H. Moore, University of Wisconsin, Milwaukee.
"A Sabermetric Model for Predicting the 1985 Al East Baseball Standings and Analysis of Results from 1984," by Steve Krevisky, University of Wisconsin, Washington County.
"Zero-Player Games: A Look at Cellular Automata," by Mike Slattery, Marquette University.
"On Arbitrary Functions of Matrices with a Unique and Closed Representation with Applications," by N.L. Petrakopoulos, University of Wisconsin, Green Bay.
"The Optimal Number of Traces of the Product of Simplices," by Hasien S. Moghadam, University of Wisconsin, Oshkosh.
"Inflating Balloons--Deflating a Misunderstanding," by Nicholas Passell, University of Wisconsin, Eau Claire.
"Some Things the Numerical Analysis Texts Forget to Mention," by John Oman, University of Wisconsin, Oshkosh.
"On F-structure Manifolds," by Lovejoy Das, University of Wisconsin, Sheboygan.
"Finite State Automata: Linking Theory and Application," by Thomas L. Naps, Lawrence University.
"Periodic Functions on the Apple Without Trigonometric Functions," by Edward Gade, III, University of Wisconsin, Oshkosh.
"A Nonparametric Test for Comparing Two Multivariate Populations," by John Bachhuber, University of Wisconsin, Oshkosh.
"Stirling's Formula and Sums of Powers of Integers," by Eric Key, University of Wisconsin, Milwaukee.
"Subgroups Satisfying a Certain Distributive Law," by Roger Erickson, University of Wisconsin, LaCrosse.
"Building Fundamental Statistical Concepts: Tukey's Exploratory Data Analysis and Its Impact on the Secondary School Mathematics Curriculum," by Gail Burrill and Henry Kepner, Whitnall High School and University of Wisconsin, Milwaukee.
"Mathematics Education in Third World Countries," by Virginia Pillsbury, Milwaukee Area Technical College.
"A More or Less Simple Approach to Limits," by Richard Menzel, University of Wisconsin, Oshkosh.
"An HGR Calculus Tutorial," by Michael Lucas, Cardinal Stritch College.
"Conic Sections in Taxicab Geometry," by Barbara E. Reynolds, Cardinal Stritch College.
"Free Subgroups in Linear Groups over Some Skew Fields," by Alexander Lichtman, University of Wisconsin, Parkside.
"An Algorithm for Minimizing a Non-linear Objective Function," by Tzu-Cheq Kao, University of Wisconsin, Oshkosh.
"An Actuarially Useful Implementation of the Maximum Likelihood Principle," by Nasser Hadidi, University of Wisconsin, Stout.
"Numerical Simulation of Fluid Flow in a Heated Enclosed Cavity," by William Shay, University of Wisconsin, Green Bay.

Panel Discussion:

"Curriculum Planning in Mathematics" (a preliminary report of the Mathematics Curriculum Task Force writing for the new Guide to Curriculum Planning in Mathematics to be issued by the Wisconsin Department of Public Instruction), by Paul Campbell (Moderator), Beloit College; LeRoy Dalton, Wauwatosa West High School; Robert Hall, University of Wisconsin, Milwaukee; John Moyer, Marquette University; Virginia Pillsbury, Milwaukee Area Technical College.

Student Presentations:

"Complex Numbers and Synthetic Geometry," by Susan Kelly, University of Wisconsin, Eau Claire.
"Taxicab Geometry," by Lou Mertens, Cardinal Stritch College.
"On the Simultaneous Numerical Computation of the Characteristic Polynomial, Adjoint Matrix, Inverse Matrix, Determinant, and the Complex Characteristic Values Associated with an Arbitrary Nonsingular Matrix," by Dana M. Johnson, University of Wisconsin, Green Bay.
"Infinitesimals Within Standard Analysis," by Mark Hopkins, University of Wisconsin, Milwaukee.

$F(\phi)M = 0$, it follows that

$$F(\psi)M \equiv 0 \bmod (\phi - \psi)M.$$

The theorem above is thus an immediate consequence of this observation and Theorem 1.1.

**References**

1. M. Braun, Differential Equations and Their Applications, 3rd Ed., Springer-Verlag, 1983.
2. J. M. Cohen, A spectral sequence automorphism theorem; applications to fibre spaces and stable homotopy, Topology, 7 (1968) 173–177.
3. ______, Clarification to a result, Topology, 9 (1970) 299–300.
4. Marvin Jay Greenberg, Note on the Cayley-Hamilton Theorem, this MONTHLY, 91 (1984) 193–195.
5. Peter Hilton, A note on finitary automorphisms, Aspects of Mathematics and its Applications, North Holland, 1985 (to appear).
6. Peter Hilton and Joseph Roitberg, On pseudo-identities I, Archiv der Mathematik, 41 (1983) 204–214.
7. Kenneth Hoffman and Roy Kunze, Linear Algebra, 2nd ed., Prentice-Hall.
8. Saunders Mac Lane and Garrett Birkhoff, Algebra, 2nd Ed., Macmillan, 1979.
9. Maurice Orzech, Onto endomorphisms are isomorphisms, this MONTHLY, 78 (1971) 357–362.
10. Joseph Roitberg, Finitary automorphisms and integral homology, Topological Topics, London Mathematical Society Lecture Notes, 86 (1983) 164–168.
11. Gilbert Strang, Linear Algebra and its Applications, 2nd Ed., Academic Press, 1980.

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## GENERALIZED LEAST SQUARES AND EIGENVALUES

ALAN SHUCHAT

*Department of Mathematics, Wellesley College, Wellesley, MA 02181*

In a recent note in this section of the MONTHLY [**2**], Hung C. Li discusses the generalized least squares problem of finding the best-fitting line $L$ to a set of points in the plane. Here "generalized" means that the sum $S$ of squares of the *perpendicular* distances from the points to $L$ should be minimized, as opposed to ordinary least squares where distances are measured vertically or horizontally. This forms the basis of orthogonal regression [**3**] and principal component analysis [**4**]. Li's approach is to treat $S$ as a function of the two parameters that determine $L$ and to minimize it by partial differentiation. However, we can obtain additional geometric insight and extend the solution more easily to higher dimensions by using linear algebra, as in [**4**]. The presentation given here is simpler and accessible to undergraduates.

If $L$ is a line and $x$ a point in $R^n$, we let $d(x, L)$ denote the orthogonal distance from $x$ to $L$. Given points $x^1, \dots, x^m$ in $R^n$, we seek a line $L$ for which the sum $S = \Sigma_k d(x^k, L)^2$ is minimal. We will show that $L$ must pass through the centroid of the given points and that its direction can be given by any eigenvector for the largest eigenvalue of a certain symmetric matrix. More generally, the $d$ largest eigenvalues determine a best-fitting affine subspace of dimension $d$.

We first show that the desired line must pass through the centroid, which for convenience we assume to be at the origin, i.e., $\Sigma_k x^k = 0$. Let $L'$ be any line $x = w + tv$ in $R^n$, where $v$ is a unit vector, and let $L$ be its translation $x = tv$ to the origin. For each $x$ in $R^n$,

$$d(x, L)^2 = \|x\|^2 - (x \cdot v)^2,$$

so

$$d(x,L')^2 = d(x-w,L)^2 = d(x,L)^2 + d(w,L)^2 - 2x\cdot[w-(w\cdot v)v].$$

This implies

$$\sum_k d(x^k,L')^2 = \sum_k d(x^k,L)^2 + md(w,L)^2 \geqslant \sum_k d(x^k,L)^2.$$

So we need only consider lines $L$ through the origin.

Now we must minimize

$$(1) \qquad S = \sum_k d(x^k,L)^2 = \sum_k \|x^k\|^2 - \sum_k (x^k\cdot v)^2$$

over $\|v\| = 1$, i.e., maximize $f(v) = \sum_k (x^k\cdot v)^2$. Let $X$ be the $m\times n$ matrix with rows $x^k$ and $A$ the $n\times n$ symmetric matrix $X^tX$. Then $f(v) = \|Xv\|^2 = Av\cdot v$, so $f$ is a quadratic form.

We now apply the spectral theorem for symmetric matrices and diagonalize $A$. In fact, maximizing a quadratic form such as $f$ is one way to prove the spectral theorem [**1**]. The matrix $A$ is orthogonally similar to the diagonal matrix $D$ of its eigenvalues $\lambda_i$, i.e., $D = P^tAP$ for an orthogonal matrix $P$. We may arrange the eigenvalues, which are real, in nonincreasing order. The quadratic form corresponding to $D$ is then

$$g(u) = Du\cdot u = \sum_i \lambda_i u_i^2.$$

Since $f(v) = g(P^tv)$ and $P$ is an isometry,

$$\max\{f(v): \|v\| = 1\} = \max\{g(u): \|u\| = 1\}.$$

But $g(u)$ lies in the interval spanned by the $\lambda_i$, so its maximum value is the largest eigenvalue $\lambda_1$ and occurs when $u = e_1$, the first standard basis vector. So the maximum of $f$ occurs when $v = Pe_1$, which is the first column of $P$ and a unit eigenvector for $\lambda_1$. This vector $v$ gives the direction of $L$. Any unit eigenvector $v_1$ of $A$ with eigenvalue $\lambda_1$ will give the same value $f(v_1) = \lambda_1$ for $f$, so if $\lambda_1$ is a repeated eigenvalue, then $L$ will not be uniquely determined.

Similarly, the worst-fitting line through the origin, for which $S$ is maximized, is the line for which $f$ is minimized and its direction is given by any unit eigenvector $v_n$ for the smallest eigenvalue $\lambda_n$. This is essentially Rayleigh's principle [**5**].

Moreover, we can view maximizing $f$ as finding the farthest points from the origin on an ellipsoid in $R^n$. The quadratic form $Av\cdot v$ is positive semi-definite, so each $\lambda_i \geqslant 0$. This quadratic form is positive-definite if and only if the vectors $x^k$ span $R^n$. If they do (as in the example below), then each $\lambda_i > 0$ and $A$ has a positive-definite symmetric square root $\sqrt{A}$, which can be constructed as follows: Let $\sqrt{D}$ be the diagonal matrix with diagonal entries $\sqrt{\lambda_i}$ and let $\sqrt{A} = P\sqrt{D}\,P^t$, so that $(\sqrt{A})^2 = A$. Then $f(v) = \|\sqrt{A}\,v\|^2$ and the image of the unit circle $\|v\| = 1$ under $\sqrt{A}$ is the ellipsoid $A^{-1}x\cdot x = 1$, where $x = \sqrt{A}\,v$ ($A^{-1}$ is also positive-definite and symmetric). By diagonalizing $A^{-1}$, $D^{-1} = P^tA^{-1}P$, we see that our problem amounts to maximizing $\|y\|$ such that $\sum_i \lambda_i^{-1}y_i^2 = 1$, where $y = P^tx$. Since $\sum y_i^2 \leqslant \lambda_1$ for all such $y$, the maximum occurs at $y = \sqrt{\lambda_1}\,e_1$. So the maximum on the ellipsoid occurs where $x = \sqrt{\lambda_1}\,v_1$, i.e., the best-fitting line lies along the major axis of the ellipsoid, in the direction of $v_1$, and the semi-major axis has length $\sqrt{\lambda_1}$. Similarly, the worst-fitting line lies along the minor axis, in the direction of $v_n$, and the semi-minor axis has length $\sqrt{\lambda_n}$.

Finally, in (1), the sum $\sum_k \|x^k\|^2$ is just the trace $\operatorname{tr} A$ of $A$, the sum of the diagonal entries of $A$. Since the trace is also the sum of the eigenvalues, the value of $S$ for the best line is the sum of the eigenvalues except for $\lambda_1$, while the value for the worst line is the sum except for $\lambda_n$.

We illustrate these ideas with the example given by Li [2]. We wish to fit a line to the seven points $(0,-2)$, $(4,-1)$, $(6,0)$, $(6,1)$, $(4,2)$, $(0,3)$, and $(-6,4)$ in $R^2$. We first translate these points so that the centroid, now at $(2,1)$, moves to the origin. This gives

$$X=\begin{pmatrix}-2 & 2 & 4 & 4 & 2 & -2 & -8\\ -3 & -2 & -1 & 0 & 1 & 2 & 3\end{pmatrix}^t,$$

and

$$A=\begin{pmatrix}112 & -28\\ -28 & 28\end{pmatrix}.$$

The eigenvalues of $A$ are

$$\lambda_1 = 70+14\sqrt{13} \approx 120.48$$

and

$$\lambda_2 = 70-14\sqrt{13} \approx 19.52.$$

Corresponding unit eigenvectors are

$$v_1 = (3+\sqrt{13},-2)/\sqrt{26+6\sqrt{13}} \approx (.96,-.29)$$

and

$$v_2 = -(3-\sqrt{13},-2)/\sqrt{26-6\sqrt{13}} \approx (.29,.96),$$

which is orthogonal to $v_1$.

The matrix $A$ is positive-definite, and $A^{-1}x\cdot x=1$ describes an ellipse in $R^2$ with equation

$$x_1^2+2x_1x_2+4x_2^2=84.$$

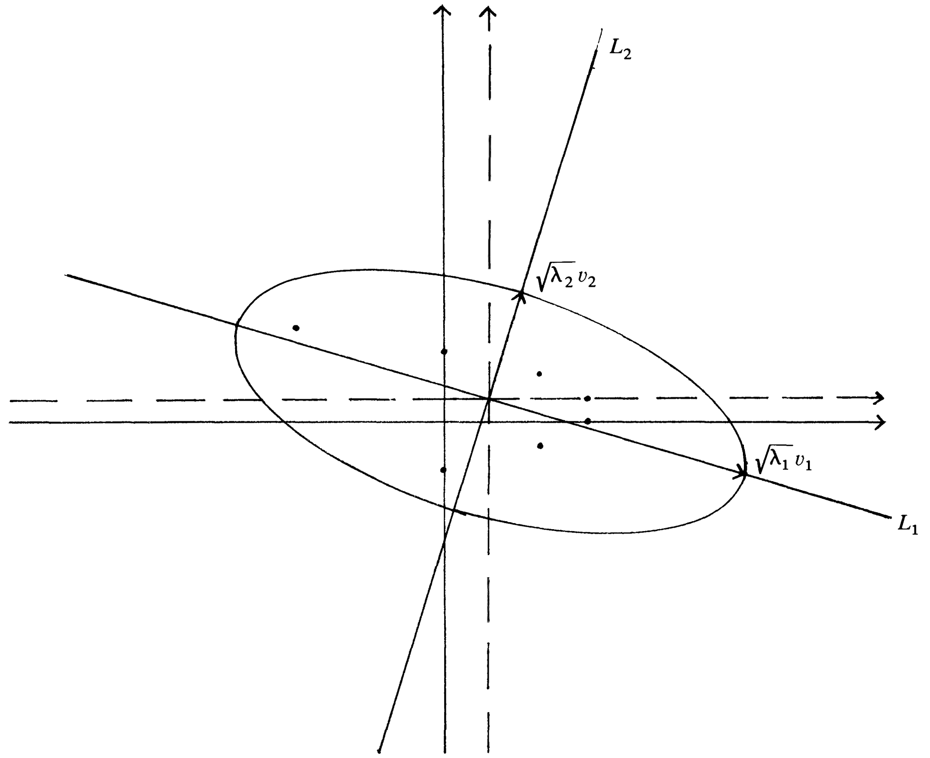



FIG. 1

Its axes lie along $v_1$ and $v_2$, with semi-axis lengths $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$, respectively. After translation back to the original centroid $(2,1)$, the best-fitting line $L_1$ becomes

$$x = (2,1) + t(3 + \sqrt{13}, -2)$$

and the worst-fitting line $L_2$ becomes

$$x = (2,1) + t(3 - \sqrt{13}, -2),$$

which are just the lines found by Li in [2]. The sum $S$ is $\lambda_2$ for $L_1$ and $\lambda_1$ for $L_2$, which explains the numerical values found by Li. Fig. 1 summarizes the example.

**Acknowledgment.** I would like to thank the MIT Operations Research Center, where I was on leave while preparing this note, for its hospitality.

### References

1. G. Birkhoff and S. Mac Lane, A Survey of Modern Algebra, Macmillan, New York, 1953.
2. H. C. Li, A generalized problem of least squares, this MONTHLY, 91 (1984) 135–137.
3. E. Malinvaud, Statistical Methods of Econometrics, 3rd ed., Elsevier, New York, 1980.
4. C. R. Rao, The use and interpretation of principal component analysis in applied research, Sankhyā, Ser. A., 26 (1964) 329–358.
5. G. Strang, Linear Algebra and Its Applications, 2nd ed., Academic Press, New York, 1980.

## FIXED POINTS AND EXAM TAKING STRATEGY

THOMAS O'BRIEN
*Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403*

In certain mathematics exams, a student is given a large number of problems and a limited time in which to answer correctly a number sufficient to achieve a passing score. A good example is provided by the early examinations in the examination program for attaining fellowship in the Society of Actuaries. The student's problem is, "How much time should be spent on a question before giving up and moving on to the next?" Our answer is, "If it is going to take longer than average, move on to the next question." In this note we explain what this answer means, why it is so, and how the question arose.

In an article in a recent newsletter of the Society of Actuaries [1], Chan provided the following statistical model for solving the above problem.

**Chan model.** Each problem is screened for $S$ minutes. After screening, the working time $W$ is known. Total time is then given by $T = S + W$. Here, $S$ and $W$ are random variables and they are assumed to have exponential distributions with $E(S) = 1/\lambda$ and $E(W) = 1/\beta$. Thus $S$ and $W$ have densities $f(x) = \lambda e^{-\lambda x}$ and $g(x) = \beta e^{-\beta x}$, $x > 0$. Further,

$$E(T) = \frac{1}{\lambda} + \frac{1}{\beta}.$$

The strategy is then: Screen each problem to determine $W$; if $W > \tau$, go on to the next problem. We seek the best moving-on time, $\tau$. The expected working time for a problem worked becomes

$$E(W|W \leqslant \tau) = \frac{\int_0^\tau x\beta e^{-\beta x}\,dx}{\int_0^\tau \beta e^{-\beta x}\,dx} = \frac{1}{\beta}\left[1 - \frac{\beta\tau e^{-\beta\tau}}{1 - e^{-\beta\tau}}\right].$$

Not every question screened will be worked out and thus the expected screening time per question worked out will increase to

$$\frac{1}{\lambda}\left[\frac{1}{1 - e^{-\beta\tau}}\right].$$

The $\tau$ we seek is that which minimizes the total time per worked question,

$$\phi(\tau)=\frac{1}{\lambda}\left[\frac{1}{1-e^{-\beta\tau}}\right]+\frac{1}{\beta}\left[1-\frac{\beta\tau e^{-\beta\tau}}{1-e^{-\beta\tau}}\right].$$

This function is easily seen to have a unique critical point at which its minimum value is attained. At the critical point,

$$\frac{1}{\lambda}+\frac{1}{\beta}=\tau+\frac{e^{-\beta\tau}}{\beta}.$$

In [**1**] the $\beta$ in the denominator of the exponential term was left out. We thus undertook to redo the computations of Chan's note and in solving for the critical point using Newton's method discovered that the minimum occurs at a fixed point of $\phi$. Indeed, solving the last equation for $e^{-\beta\tau}$ and substituting into the expression for $\phi(\tau)$, we get $\phi(\tau)=\tau$. We then inquire whether this is merely an idiosyncrasy of the chosen model.

**Alternative model.** Here, $S$, $W$ and $T$ are as above but the densities of $S$ and $W$ are respectively

$$f(x)=x\lambda^2e^{-\lambda x}\quad\text{and}\quad g(x)=x\beta^2e^{-\beta x}.$$

Then,

$$E(S)=\frac{2}{\lambda}\quad\text{and}\quad E(W)=\frac{2}{\beta}.$$

In this model, the modes of the distributions are at $1/\lambda$ and $1/\beta$ rather than at 0 as in the Chan model. Calculating as in the first model we obtain

$$E(W|W\leqslant\tau)=\frac{2}{\beta}\left[1-\frac{\tau^2\beta^2e^{-\beta\tau}}{2(1-\tau\beta e^{-\beta\tau}-e^{-\beta\tau})}\right]$$

and the expected screening time per problem worked becomes

$$\frac{2}{\lambda}\left[\frac{1}{1-\tau\beta e^{-\beta\tau}-e^{-\beta\tau}}\right].$$

Summing the last two expressions yields

$$\phi(\tau)=\frac{2}{\beta}+\frac{2-\tau^2\lambda\beta e^{-\beta\tau}}{\lambda(1-\tau\beta e^{-\beta\tau}-e^{-\beta\tau})}.$$

Again, there is a unique critical point yielding a minimum and it is the solution to the equation

$$\tau+\left(\tau+\frac{2}{\beta}\right)e^{-\beta\tau}=\frac{2}{\beta}+\frac{2}{\lambda}.$$

Solving for $e^{-\beta\tau}$ and substituting into the expression for $\phi$, we again obtain $\phi(\tau)=\tau$ and we are led to consider a general explanation for this phenomenon.

**General model.** Suppose that $S$ and $W$, again with the same meaning as above, have distributions with continuous positive valued densities $f$ and $g$, respectively, defined on $(0,\infty)$ and that

$$E(S)=\gamma\quad\text{and}\quad E(W)=\alpha$$

$$E(W|W\leqslant\tau)=\frac{\int_0^\tau xg(x)\,dx}{\int_0^\tau g(x)\,dx}.$$

The screening time per question worked becomes $\dfrac{\gamma}{\int_0^\tau g(x)\,dx}$. Thus

$$\phi(\tau) = \frac{\gamma + \int_0^\tau xg(x)\,dx}{\int_0^\tau g(x)\,dx},$$

$$\phi'(\tau) = \frac{\tau g(\tau)\int_0^\tau g(x)\,dx - g(\tau)\left(\gamma + \int_0^\tau xg(x)\,dx\right)}{\left[\int_0^\tau g(x)\,dx\right]^2}$$

$$= (\tau g(\tau) - \phi(\tau)g(\tau))\Big/\int_0^\tau g(x)\,dx$$

$$\phi'(\tau) = (-g(\tau)(\phi(\tau) - \tau))\Big/\int_0^\tau g(x)\,dx.$$

Since $\gamma > 0$, $\lim_{\tau \to 0}\phi(\tau) = +\infty$ and $\lim_{\tau \to \infty}\phi(\tau) = \alpha + \gamma$. Also the graph of $\phi$ cannot cross the diagonal from below. It follows that $\phi$ has a unique fixed point and it occurs at the min of $\phi$.

We now restate our answer to the question posed at the beginning. If, after screening, the estimate of the time to work the problem is longer than the average time spent per problem worked, then move on to the next problem. This answer may not be terribly helpful to the test taker and so we present in a table the results of the numerical calculations for the two models above for a 60 question three-hour exam with $E(T) = 6$.

| | $E(S)$ | $E(W)$ | Best $\tau$ | Answers | Screens |
|---|---|---|---|---|---|
| Chan model | 1 | 5 | 3.534 | 50.9 | 100.57 |
| | 1.77 | 4.23 | 4.561 | 39.47 | 60 |
| | 2 | 4 | 4.793 | 37.5 | 53.83 |
| | 3 | 3 | 5.524 | 32.6 | 38.71 |
| | 4 | 2 | 5.895 | 30.5 | 32.24 |
| Alternative model | 1 | 5 | 4.367 | 41.2 | 79.13 |
| | 1.4 | 4.6 | 4.855 | 37.1 | 60 |
| | 2 | 4 | 5.358 | 33.6 | 44.9 |
| | 3 | 3 | 5.818 | 30.9 | 34.45 |
| | 4 | 2 | 5.980 | 30.1 | 30.64 |

Thus, the first model suggests that, on this test, a problem should be screened for 1.77 minutes and then if it is estimated that the time needed to work the problem is more than 4.561 minutes, go on to the next. The second suggests screening for 1.4 minutes with a moving-on time of 4.855 minutes.

**Reference**

1. Beda Chan, Moving On, The Actuary, vol. 18, no. 1, Jan. 1984, 4–5.

**ANSWER TO PHOTO ON PAGE 630**

Alfred Haar (1885–1933).

## A COMBINATORIAL PROOF OF A CRITERION FOR NORMALITY

JAMES C. OWINGS, JR.
*Department of Mathematics, University of Maryland, College Park, MD 20742*

It is well known that if $G$ is a finite group, $p$ is the smallest prime that divides $|G|$, and $H$ is a subgroup of $G$ of index $p$, then $H$ is normal in $G$. This is usually proved by representing $G$ as a group of permutations on the left cosets of $H$, along the lines of Herstein [**2**, pp. 72–73] (although Herstein does not explicitly mention this result). We present here a simple theorem relating right and left cosets from which the results easily follows. We also discuss a related theorem of Marshall Hall, Jr.

THEOREM. *If $H$ and $K$ are subgroups of a group $G$, $L$ is a left coset of $H$, and $R_1$, $R_2$ are right cosets of $K$ such that $L \cap R_1 \neq \varnothing$ and $L \cap R_2 \neq \varnothing$, then $L \cap R_1$ has the same cardinality as $L \cap R_2$.*

*Proof.* Let $w_1 \in L \cap R_1$, $w_2 \in L \cap R_2$. Then

$$L = w_1H = w_2H,\ R_1 = Kw_1, \text{ and } R_2 = Kw_2.$$

So

$$L \cap R_1 = w_1H \cap Kw_1 = \left(w_1Hw_1^{-1} \cap K\right)w_1$$

and

$$L \cap R_2 = w_2H \cap Kw_2 = \left(w_2Hw_2^{-1} \cap K\right)w_2.$$

But since $w_1H = w_2H$,

$$w_1Hw_1^{-1} = w_2Hw_1^{-1} = w_2(w_1H)^{-1} = w_2(w_2H)^{-1} = w_2Hw_2^{-1}.$$

Hence, both $L \cap R_1$ and $L \cap R_2$ are right cosets of $w_1Hw_1^{-1} \cap K$, a subgroup of $G$. This proves the theorem.

COROLLARY. *Let $G$ be a finite group, let $p$ be the smallest prime that divides $|G|$, and let $H$ be a subgroup of $G$ of index $p$. Then $H$ is normal in $G$.*

*Proof.* Let $g \in G \setminus H$; then $gH \cap H = \varnothing$. Hence, if $k$ is the number of right cosets of $H$ that $gH$ intersects, $k \leqslant p - 1$. By our theorem, the size $|gH \cap R_i|$ of these intersections is constant, say $c$. Since

$$|H| = |gH| = \sum_{gH \cap R_i \neq \varnothing} |gH \cap R_i| = ck,$$

$k$ is a divisor of $|H|$ and, by Lagrange's Theorem, a divisor of $|G|$. Therefore $k = 1$. So $gH = Hg$.

Suppose $H$ and $K$ are subgroups of $G$. Marshall Hall, Jr. [**1**, p. 14, Lemma 1.7.1] has shown that if $L_1$ and $L_2$ are left cosets of $H$, then the sets

$$\{R \mid R \text{ is a right coset of } K \text{ and } L_1 \cap R \neq \varnothing\}$$

and

$$\{R \mid R \text{ is a right coset of } K \text{ and } L_2 \cap R \neq \varnothing\}$$

are either equal or disjoint. Obviously, both our result and Hall's result remain true if the words "right" and "left" are interchanged. By putting the two results together, we get a nice picture of the interaction between right and left cosets. The left cosets of $H$ and the right cosets of $K$ can be partitioned into a finite number of blocks $B_1, \dots, B_k$ and $B_1', \dots, B_k'$, respectively, such that if $L \in B_i$ and $R \in B_j'$, then $L \cap R \neq \varnothing$ if and only if $i = j$. Our theorem then shows that, if

$B_i = \{L_1, \ldots, L_s\}$ and $B_i' = \{R_1, \ldots, R_t\}$, then all the intersections $L_i \cap R_j$, $1 \leqslant i \leqslant s$, $1 \leqslant j \leqslant t$, have the same size.

**References**

1. Marshall Hall, Jr., The Theory of Groups, Macmillan, New York, 1959.
2. I. N. Herstein, Topics in Algebra, Xerox Publishing Company, New York, 1975.

---

# THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## ON THE DIFFERENTIABILITY OF FUNCTIONS OF SEVERAL VARIABLES

MICHAEL W. BOTSKO AND RICHARD A. GOSSER
*Department of Mathematics, Saint Vincent College, Latrobe, PA 15650*

Let $f$ be a real-valued function defined on $D$ a neighborhood of $a$ in $R^n$. Then $f$ is said to be differentiable at $a$ if there is $b \in R^n$ and $\eta$: $D \to R$ such that for all $x \in D$

$$(1) \qquad f(x) - f(a) = b \cdot (x - a) + \eta(x)\|x - a\| \quad \text{and} \quad \eta(x) \to 0 \text{ as } x \to a,$$

where $a \cdot b$ is scalar product.

The purpose of this note is to present an equivalent condition for differentiability of functions of several variables, which is often easier to apply than the definition itself. The condition is similar to the one used by Carathéodory for functions of one variable, see [**1**]. Our main result is the following:

THEOREM. *Let $f$: $D \to R$ where $D \subset R^n$ is a neighborhood of $a$. Then $f$ is differentiable at $a$ if and only if there exists $h$: $D \to R^n$ continuous at $a$ and satisfying*

$$(2) \qquad f(x) - f(a) = h(x) \cdot (x - a) \quad \textit{for all } x \textit{ in } D.$$

*Proof.* Suppose (1) holds, and for $i = 1, 2, \ldots, n$, let $a_i$, $x_i$, and $b_i$ denote the $i$th component of $a$, $x$ and $b$. Also for each $i$ define

$$h_i(x) = \begin{cases} b_i + \eta(x)\dfrac{x_i - a_i}{\|x - a\|} & \text{if } x \neq a, \\ b_i & \text{if } x = a. \end{cases}$$

Clearly each $h_i$: $D \to R$ is continuous at $a$, and for $x \neq a$

$$\begin{aligned} f(x) - f(a) &= b \cdot (x - a) + \eta(x)\|x - a\| \\ &= \sum_i b_i(x_i - a_i) + \frac{\eta(x)}{\|x - a\|}\sum_i (x_i - a_i)^2 \\ &= \sum_i \left(b_i + \frac{\eta(x)(x_i - a_i)}{\|x - a\|}\right)(x_i - a_i) \\ &= \sum_i h_i(x)(x_i - a_i) \end{aligned}$$

$$= h(x) \cdot (x - a),$$

where $h(x) = (h_1(x), h_2(x), \ldots, h_n(x))$ and $h$ is continuous at $a$. Since the equality obviously holds for $x = a$, we have

$$f(x) - f(a) = h(x) \cdot (x - a) \quad \text{for all } x \text{ in } D.$$

Hence (1) $\Rightarrow$ (2).

Conversely, suppose there exists $h$: $D \to R^n$ continuous at $a$ and satisfying (2). For each $i = 1, 2, \ldots, n$, define $\eta_i$: $D \to R$ by

$$\eta_i(x) = h_i(x) - h_i(a)$$

and $\gamma$: $D \to R^n$ by

$$\gamma(x) = (\eta_1(x), \eta_2(x), \ldots, \eta_n(x)).$$

Hence

$$f(x) - f(a) = h(x) \cdot (x - a) = (h(a) + \gamma(x)) \cdot (x - a).$$

Put $b = h(a)$ and define $\eta$ so that,

$$\eta(x) = \begin{cases} \dfrac{\gamma(x) \cdot (x - a)}{\|x - a\|} & \text{if } x \neq a, \\ 0 & \text{if } x = a. \end{cases}$$

For all $x \in D$ we have

$$\begin{aligned} f(x) - f(a) &= b \cdot (x - a) + \gamma(x) \cdot (x - a) \\ &= b \cdot (x - a) + \eta(x) \|x - a\|. \end{aligned}$$

Since $h(x) \to h(a)$ as $x \to a$, $\eta(x) \to 0$ as $x \to a$ and the proof is complete.

If $f$ is differentiable at $a$, it is easy to show that $h(a)$ is the gradient vector of $f$ at $a$.

As an example of using our criterion, let us show that the following function is differentiable at $\mathbf{0}$. Define $f$: $R^n \to R$ such that

$$f(x) = \begin{cases} \sum_i x_i^{a_i} \sin \dfrac{1}{\|x\|} & \text{if } x \neq \mathbf{0}, \\ 0 & \text{if } x = \mathbf{0}, \end{cases}$$

where $a_i > 1$ for each $i$. Now define $h$: $R^n \to R^n$ so that

$$h(x) = \begin{cases} \sum_i \left( x_i^{b_i} \sin \dfrac{1}{\|x\|} \right) e_i & \text{if } x \neq \mathbf{0}, \\ (0, 0, \ldots, 0) & \text{if } x = \mathbf{0}, \end{cases}$$

where $b_i = a_i - 1$ and $e_i$ is the unit vector with 1 in the $i$th position and 0 elsewhere. Obviously $f(x) = h(x) \cdot x$ for all $x$ in $R^n$ and $h$ is continuous at $\mathbf{0}$. Thus $f$ is differentiable at $\mathbf{0}$.

It is easy to prove that sums, products and quotients of differentiable functions are differentiable using the criterion in this paper. In addition the usual chain rule and directional derivative formulas are readily established using this technique. The reader is invited to work through some of these or even to formulate others.

As an example we give a short proof of the following well-known result.

THEOREM. *Let $f$ and $g$ be real-valued functions defined on a neighborhood of $a$ in $R^n$. If $f$ is differentiable at $a$, $f(a) = 0$, and $g$ is continuous at $a$, then $fg$ is differentiable at $a$.*

*Proof.* Since $f$ is differentiable at $a$, there exists $D$ a neighborhood of $a$ and $h$: $D \to R^n$

continuous at $a$ such that

$$f(x) = h(x) \cdot (x - a) \quad \text{for all } x \text{ in } D.$$

Obviously $D$ can be chosen so that $g$ is defined on $D$. Multiplying by $g(x)$ yields

$$f(x) g(x) = g(x) h(x) \cdot (x - a) \quad \text{for all } x \text{ in } D.$$

Since $gh\colon D \to R^n$ which is clearly continuous at $a$, it follows at once that $fg$ is differentiable at $a$.

**Reference**

1. Maurice Heins, Complex Function Theory, Academic Press, New York (1968) 58.

---

# PROBLEMS AND SOLUTIONS



*Send all* ***proposed*** *problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions*** *should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by March 31, 1986. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3111. *Proposed by Themistocles M. Rassias, Athens, Greece.*

Prove that

$$\frac{\pi}{6} < \int_0^1 \frac{dx}{\sqrt{4 - x^2 - x^3}} < \frac{\pi\sqrt{2}}{8}.$$

E 3112. *Proposed by M. J. Pelling, London, England.*

Let $Q_1$ be the set of rationals greater than 1. Does $Q_1$ contain a minimal set $B$ such that every $q \in Q_1$ is a product, not necessarily unique, of elements in $B$?

E 3113. *Proposed by Frank Schwellinger (student), Karlsruhe, West Germany.*

Prove that there is a continuous, strictly decreasing function $g(x)$ of the real line $R$ onto $R$ such that $g(g(x)) = 2x + 2$, but that there is no such function satisfying $g(g(x)) = x + 1$.

E 3114. *Proposed by M. J. Pelling, London, England.*

Find the largest cube that can be inscribed in some tetrahedron of volume 1.

E 3115. *Proposed by I. J. Schoenberg, University of Wisconsin-Madison.*

(a) Let $z_j$ $(j = 1,\ldots,n)$ be the zeros of the polynomial

$$P_n(z) = z^n + a_2 z^{n-2} + \cdots + a_n \tag{1}$$

with complex coefficients. (Note $\Sigma z_j = 0$.) Show that

$$\sum_{j=1}^{n} |z_j|^2 \geqslant 2|a_2|,$$

with the equality sign if and only if the $z_j$ all lie on a straight line passing through 0 in the complex plane.

(b)* Let $w_j$ $(j = 1,\ldots,n-1)$ be the zeros of the polynomial $P_n'(z)$, from (1). Show that

$$\sum_{j=1}^{n-1} |w_j|^2 \leqslant \frac{n-2}{n} \sum_{j=1}^{n} |z_j|^2,$$

with equality if and only if all the $z_j$ lie on a straight line passing through 0 in the complex plane.

E 3116. *Proposed by M. Laub, Jerusalem, Israel.*

(a) If $x$ and $y$ are nonnegative real numbers, prove that $x^x + y^y \geqslant x^y + y^x$.

(b) If $\{x_1, x_2, \ldots, x_n\}$ is a sequence of nonnegative real numbers and $\{y_1, y_2, \ldots, y_n\}$ is any permutation of this sequence, prove or disprove that

$$x_1^{x_1} + x_2^{x_2} + \cdots + x_n^{x_n} \geqslant x_1^{y_1} + x_2^{y_2} + \cdots + x_n^{y_n}.$$

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Properties of the Matrix of a Rosette

E 2898 [1981, 538]. *Proposed by U. Abel, Heidelberg DKFZ, West Germany.*

A rosette is a directed graph $G$: $(V \equiv \{x_0,\ldots,x_n\}, E)$ with nonzero arc-weights, consisting of cycles which all pass through the central point $x_0$.

Let $A$ be the matrix corresponding to $G$, i.e., $a_{ij}$ = weight on $(x_i, x_j)$, and let $m_i$ and $n_i$ denote the number of cycles in $G$ of length $> i$ and $= i$, respectively. Assume that the maximal cycle length $l$ is greater than 1.

Prove the following statements:

(a) $\dim\ker A^k = \sum_{i=1}^{k} (m_i - 1),\ k < l$;

(b) $\dim\ker A^l = \dim\ker A^{l-1} = 1 - l + \sum_{k=2}^{l} n_k(k-1)$.

*Solution by the proposer.* Inductively define

$M_1 = \{x \in V | x \neq x_0$ and $(x, x_0)$ is an edge of $G\}$;

Thus

$$\sum_{p=1}^{l-1} m_p = (l-1)n_l + (l-2)n_{l-1} + \cdots + n_2$$
$$= \sum_{p=2}^{l} n_p(p-1).$$

Therefore

$$(5) \qquad \sum_{p=1}^{l-1} (m_p - 1) = (1-l) + \sum_{p=1}^{l-1} n_p(p-1).$$

Together, (a), (4) and (5) yield (b).

### An Application of Fubini's Theorem

E 2965 [1982, 594]. *Proposed by Bu Qi-yue, Shanghai Jiao-tung University, People's Republic of China.*

Let $0 < h, A \subseteq [a, b]$, $A$ measurable. Prove

$$\frac{1}{2h}\int_a^b m[A \cap (x-h, x+h)]\, dx \leqslant mA.$$

*Solution by K. F. Andersen, University of Alberta.* If $\chi_A(t)$ denotes the characteristic function of $A$, then we have

$$\int_a^b m[A \cap (x-h, x+h)]\, dx = \int_a^b dx \int_{|x-t|<h} \chi_A(t)\, dt.$$

Applying Fubini's Theorem to the double integral on the right, we see that this is clearly less than or equal to

$$\int_{-\infty}^{\infty} \chi_A(t)\, dt \int_{|x-t|<h} dx = (mA)(2h),$$

as required.

Also solved by 26 other readers and the proposer.

### An Irreducibility Criterion

E 2969 [1982, 697]. *Proposed by John Brillhart, The University of Arizona, and Constantin Sevici, University of Michigan.*

Let $f(x) = a_0x^n + \cdots + pa_n$ be a polynomial with integer coefficients such that $a_0a_n \neq 0, (a_0, a_1, \ldots, pa_n) = 1$, and let $p$ be a prime such that $p > \sum_{s=0}^{n-1}|a_s||a_n|^{n-1-s}$. Prove that $f(x)$ is irreducible in $Z[x]$.

*Editorial note.* The original problem was slightly misstated, as the example (due to W. Janous, Austria) $n = 2, a_0 = 1, a_1 = 2, a_2 = 1/p, p$ any odd prime, shows. The intention was to assume that $a_n$, not merely $pa_n$, is an integer. With this amendment we have the following.

*Solution by E. R. Gentile, Argentina.* Let $\alpha \in C$ be a root of $f(x)$. We claim that $|\alpha| > |a_n|$. Otherwise, $f(\alpha) = 0$ implies

$$|pa_n| = |a_0\alpha^n + \cdots + a_{n-1}\alpha| \leqslant |a_n| \sum_{s=0}^{n-1} |a_s||a_n|^{n-1-s} < |a_n||p|,$$

a contradiction. Assume that $f(x)$ is reducible with respect to $Z[x]$. Then we have in $Z[x]$:

$$f(x) = (b_0x^r + \cdots + b_rp)(c_0x^t + \cdots + c_t), \quad 1 \leqslant t \leqslant (n-1),$$

with $b_0c_0 = a_0, b_rc_t = a_n$. Let $\alpha_1, \ldots, \alpha_t$ be the roots of the second factor. We have

$$\prod_{i=1}^{t} \alpha_i = \pm \frac{c_t}{c_0}.$$

Therefore

$$|a_n| \geqslant \frac{|c_t|}{|c_0|} = \prod_{i=1}^{t} |\alpha_i| > |a_n|^t,$$

which is absurd. Hence $f(x)$ must be irreducible with respect to $Z[x]$.

Also solved by O. P. Lossers (The Netherlands), D. P. Mehendale and M. R. Modak (India), B. L. Osofsky, University of South Alabama Problem Group, and the proposers.

### Simplices, Circumspheres, and Centroids

E 2987 [1983, 133]. *Proposed by George Tsintsifas, Thessaloniki, Greece.*

Let $S_n = A_1A_2 \ldots A_{n+1}$ be an $n$-simplex in $\mathbb{R}^n$ and $M$ a point inside its circumsphere $S$: $(0, R)$. The straight line $A_iM$ intersects the sphere $(0, R)$ at the point $A_i'$. We denote $K = \sum_{i=1}^{n+1} |A_iM|/|MA_i'|$. Let $G$ be the centroid of $S_n$. Prove:

(a) $K > n+1$ if and only if $M$ lies outside the sphere $(s)$ with diameter $OG$.

(b) $K = n+1$ if and only if $M$ lies on the sphere $(s)$.

(c) $K < n+1$ if and only if $M$ lies in the interior of the sphere $(s)$.

*Solution by Leon Gerber, St. John's University, Jamaica, New York.*

$$K = \sum_{i=1}^{n+1} \frac{(A_iM)^2}{|MA_i'||A_iM|} = \sum_{i=1}^{n+1} \frac{(A_iM)^2}{R^2 - (MO)^2}.$$

For any point $P$ let $\vec{p} = \overrightarrow{OP}$ and $p = |OP|$. Then

$$K(R - m^2) = \sum_{i=1}^{n+1} \|\vec{m} - \vec{a}_i\|^2 = \sum_{i=1}^{n+1} (m^2 - 2\vec{m} \cdot \vec{a}_i + a_i^2)$$

$$= (n+1)m^2 - 2(n+1)\vec{m} \cdot \vec{g} + (n+1)R^2$$

implies

$$m^2 - \frac{2(n+1)}{K+n+1}\vec{m} \cdot \vec{g} + \left(\frac{n+1}{K+n+1}\right)^2 g^2 = R^2 - \frac{2(n+1)}{K+n+1}R^2 + \left(\frac{n+1}{K+n+1}\right)^2 OG^2 = f^2,$$

or

$$\left\| \vec{m} - \frac{n+1}{K+n+1}\vec{g} \right\|^2 = f^2.$$

That is, for a fixed value of $K$, $M$ lies on the sphere with radius $f$ and center $F$ given by

$$\vec{f} = \frac{K}{K+n+1}\vec{o} + \frac{n+1}{K+n+1}\vec{g}.$$

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by March 31, 1986. The solver's full post-office address should be on each sheet.*

6503. *Proposed by Gérard Letac, Université Paul-Sabatier, Toulouse, France.*

The Markov chain $(X_n)_{n=0}^{\infty}$ on the interval $(-1,1)$ is such that, given $X_n$, $X_{n+1}$ is uniformly distributed on $(X_n, 1)$ if $X_n \leqslant 0$, and on $(-1, X_n)$ if $X_n > 0$. Find the stationary distribution.

6504. *Proposed by Michael S. Perkins, Stanford University.*

Find each function $f(z)$ which is analytic at $z = 0$ and for which $nf_n\left(\frac{z}{n}\right)$ converges uniformly to $f(z)$ in some neighborhood of $z = 0$, where $f_1 = f$ and $f_n$ is the composition of $f$ with itself $n$ times.

## SOLUTIONS OF ADVANCED PROBLEMS

### The Constant Term in an Expansion

6458 [1984, 259]. *Proposed by Mark Haiman and David Richman, MIT.*

Determine the constant term $f(p)$ of

$$(x + y + x^{-1} + y^{-1})^p,$$

where $x$ and $y$ are noncommutative variables.

*Solution by Allen J. Schwenk, Office of Naval Research, Arlington, Virginia.* The expansion of this expression produces $4^p$ terms which we shall interpret as walks of length $p$ in an infinite tree that has four edges incident at each vertex. Each edge contributes one of the four factors $x$, $x^{-1}$, $y$, or $y^{-1}$ as it is traversed. The same edge traversed in the opposite direction contributes its inverse factor. Now a term reduces to the constant 1 if and only if it represents a closed walk, that is, it terminates at its origin. Of course this means $p = 2n$ must be even, and $f(2n)$ is the number of possible closed walks of length $2n$.

We shall use generating functions to enumerate closed walks. Let $A(z) = \sum_{n=1}^{\infty} a_{2n} z^{2n}$ denote the generating function for closed walks of length $2n$ with a prescribed first step $uv$ and returning to the origin $u$ exactly once. Now if $v$ is visited $k+1$ times before returning to $u$, between each pair of occurrences of $v$ there must be a closed walk using one of three available first steps. Thus, a closed walk returning $k+1$ times to $v$ before returning to $u$ is counted by $z^2 3^k A^k(z)$. Summing for all $k \geqslant 0$ we find

$$A(z) = z^2 \sum_{k=0}^{\infty} 3^k A^k(z) = \frac{z^2}{1 - 3A(z)}.$$

Observing that $A(0) = 0$ we conclude that

$$A(z) = \frac{1 - (1 - 12z^2)^{1/2}}{6}.$$

Now let $B(z) = \sum_{n=0}^{\infty} b_{2n} z^{2n}$ be the series for all closed walks starting at $u$ with repeated returns allowed. A walk that returns exactly $k$ times is counted by $4^k A^k(z)$. Thus,

$$B(z) = \sum_{k=0}^{\infty} 4^k A^k(z) = \frac{1}{1 - 4A(z)},$$

from which it follows that

$$B(z)=\frac{2(1-12z^2)^{1/2}-1}{1-16z^2}.$$

From this we easily derive the recurrence relation

$$b_{2n}=16b_{2n-2}-\frac{2}{2n-1}\binom{2n}{n}3^n$$

with the initial condition $b_0=1$, and from the recurrence relation we obtain the formula

$$f(2n)=b_{2n}=16^n-\sum_{k=1}^{n}\frac{2}{2k-1}\binom{2k}{k}3^k16^{n-k}.$$

This approach extends easily to handle $m$ noncommuting variables. The constant term of $(\sum_{i=1}^{m}(x_i+x_i^{-1}))^{2n}$ is found to be

$$(2m)^{2n}-\sum_{k=1}^{n}\frac{m}{2k-1}\binom{2k}{k}(2m-1)^k(2m)^{2n-2k}.$$

Also solved by L. E. Clarke (England), S. V. Kanetkar, O. P. Lossers (The Netherlands), J. G. Mauldon, R. W. K. Odoni (England), James Propp, B. E. Sagan, Lajos Takàcs, A. Tissier (France) and the proposers. Partially solved by Richard Staum and D. H. Wiedemann (Canada).

### On Infinite Values of a Derivative

6459 [1984, 314]. *Proposed by Mark Bowron (student), University of Wisconsin.*

Let $C$ be Cantor's ternary set, and let $g:[0,1]\to[0,1]$ be the Cantor-Lebesgue function. Define a measure $\mu$ on $C$ by setting $\mu(S)=mg(S)$ for $S\subseteq C$, where $m$ is Lebesgue measure. We know that $g'(x)=0$ a.e. $(m)$ in $[0,1]$. Show that $g'(x)=\infty$ a.e. $(\mu)$ in $C$.

*Solution by F. S. Cater, Portland State University, Portland Oregon.* The result is an immediate consequence of Theorem 4.6 on page 271 of S. Saks, *Theory of the Integral*, second revised edition. This theorem states that if a nondecreasing function $f$ has a finite Dini derivative at each point of a set $E$ with $mE=0$, then $mf(E)=0$.

Ivan Netuka and Jiří Veselý (Czechoslovakia) and Richard B. Tucker observed that the result follows easily from Theorem 8.10 on page 169 of W. Rudin, *Real and Complex Analysis*, second edition. Also solved by David Moews, Alexander V. Stanoyevitch, Wm. Douglas Withers, and the proposer.

### A Pathological Function

6460 [1984, 371]. *Proposed by L. Richard Duffy (student), University of Chicago.*

It is well known that a discontinuous *linear* real function (i.e., one satisfying $f(x+y)=f(x)+f(y)$, all $x,y\in\mathbb{R}$) must be unbounded on any open interval. Is there a linear function which is so pathological that it actually assumes *all* real values on any open interval?

*Solution by Detlef Laugwitz, Technische Hochschule Darmstadt, West Germany.* The answer is "yes". Let $B$ be a Hamel basis of $\mathbb{R}$ over $\mathbb{Q}$ with $1\in B$. Since $B$ has the same cardinality as $\mathbb{R}$, we may write $B=\{b_x|x\in\mathbb{R}\}$ with $b_0=1$. Define a $\mathbb{Q}$-linear function $f$ by setting $f(b_x)=x$. For each $x$ there exists a rational number $q$ such that $b_x+q$ lies in a given interval, and

$$f(b_x+q)=f(b_x)+qf(1)=f(b_x)=x.$$

Also solved by the proposer and twenty-four others. A number of solvers pointed out that a solution of the problem is contained in a paper by F. B. Jones, Connected and disconnected plane sets and the functional equation $f(x)+f(y)=f(x+y)$, *Bull. Amer. Math. Soc.*, 48 (1942), 115–120.

**An Exponential Sum Inequality**

6461 [1984, 371]. *Proposed by L. E. Mattics, University of South Alabama.*

Let $p$ be a prime and suppose $1 \leqslant h < p$ where the order of $h$ modulo $p$ is $(p-1)/\nu$. Prove that

$$(\nu - 1)\sqrt{p} + 1 \geqslant \left| \sum_{m=1}^{p-1} e^{2\pi i h^m/p} \right|.$$

*Solution by Harald Niederreiter, Austrian Academy of Sciences, Vienna, Austria.* More general results are available in the literature. Let $M \geqslant 2$, $b$, and $h$ be integers with $\gcd(b, M) = \gcd(h, M) = 1$ and with the order of $h$ modulo $M$ being $\tau = \phi(M)/\nu$, where $\phi$ is Euler's totient function. Then it is shown in [**1**, Theorem 8.3] that

$$\left| \sum_{m=1}^{\tau} e^{2\pi i b h^m/M} \right| \leqslant \sqrt{M} - \frac{1}{\nu}(\sqrt{M} - 1).$$

The desired result follows immediately by putting $M = p$, $b = 1$, and noting that modulo $p$ the powers $h^m$, $m = 1, 2, \ldots, p-1$, run through $\nu$ full periods.

**Reference**

**1.** H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.*, 84 (1978), 957–1041.

Also solved by Karl Dilcher (Canada), Ronald J. Evans, S. V. Kanetkar, O. P. Lossers (The Netherlands), R. W. K. Odoni (England), James C. Smith, and the proposer.

---

# REVIEWS

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*The Lore of Prime Numbers*. By George P. Loweke. Vantage Press, New York, 1982. vii + 259 pp. \$17.95.

CLAUDIA SPIRO
*Mathematics Department, State University of New York at Buffalo, Buffalo, NY 14214*

The ancient Greeks defined a prime as an integer $p$ greater than 1, which cannot be factored into a pair of integers, both of which exceed 1. Integers greater than 1 which are not primes are called composite numbers. Every integer $n > 1$ can be written as a product of prime numbers, and this representation of $n$ is unique, except for the order of the factors. Thus, the primes freely generate the semigroup of positive integers under multiplication. This result, which was known to Euclid, is called the Fundamental Theorem of Arithmetic.

Euclid also knew that there are infinitely many primes; however, there is still no known "simple rule" for finding the $n$th prime. For example, it is natural to consider a nonconstant polynomial in $n$ with integer coefficients. But, it can readily be shown (see Theorem 21 on p. 18 of [**9**]) that any such polynomial takes infinitely many composite values. As an interesting and well-known example (see pp. 106–7 of [**16**]), we remark that $m^2 - m + 41$ is prime for $m = 1, 2, \ldots, 40$, but that 41 evenly divides $m^2 - m + 41$ whenever $m$ is an integer multiple of 41. It is

important to mention that there are nonconstant polynomials with integer coefficients in *several* variables such that when they are evaluated at positive integers, and the resulting value is positive, it is prime. For a good, accessible account of these results, we refer the reader to [**5**]. It is also known [**4**] that no nonconstant rational function of one variable, with real coefficients, can assume only prime values at the integers. W. H. Mills [**17**] has shown the existence of a real number $A > 0$ such that $[A^{3^n}]$ is prime for all positive integers $n$, where $[x]$ denotes the greatest integer not exceeding $x$. His proof relies on the result of Ingham [**11**] that there is a positive constant $c_1$ such that for all real $x \geqslant 1$, there is a prime between $x$ and $x + c_1 x^{5/8}$. Unfortunately, to find the $A$ in Mills' result (or at least to approximate it), the very primes that his rule gives have to be found first.

Although it is not easy to predict the terms of the sequence of primes, this sequence behaves well on the average. To indicate what we mean by "on the average", we first write $\pi(x)$ for the number of primes not exceeding the positive real number $x$. In addition, if $f(x), g(x)$ are real-valued functions of $x$, we write $f(x) = o(g(x))$ to mean that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

On the basis of numerical tables, Carl Friedrich Gauss conjectured that

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1, \tag{1}$$

where $\log x$ signifies the natural logarithm of $x$. An account of his reasoning is given on pp. 4–5 of [**14**]. Equation (1), which can be restated as

$$\pi(x) - \frac{x}{\log x} = o\left(\frac{x}{\log x}\right), \tag{2}$$

has come to be called the Prime Number Theorem. It was first proved, independently, by J. Hadamard [**7**] and by Charles de la Vallée Poussin [**20**] in 1896. Their proofs required complex variable theory. In 1899, de la Vallée Poussin [**21**] improved the estimate (2) by showing that there are positive constants $c_2$ and $c_3$ such that

$$\left|\pi(x) - \int_2^x \frac{dt}{\log t}\right| < c_2 x e^{-c_3\sqrt{\log x}}, \tag{3}$$

for $x \geqslant 2$. We remark that

$$\int_2^x \frac{dt}{\log t} - \frac{x}{\log x} - \frac{x}{(\log x)^2} = o\left(\frac{x}{(\log x)^2}\right),$$

and that the right side of (3) grows more slowly than $x(\log x)^{-2}$, so that (3) implies that

$$\pi(x) - \frac{x}{\log x} - \frac{x}{(\log x)^2} = o\left(\frac{x}{(\log x)^2}\right).$$

In fact, (3) can be made to yield an asymptotic expansion for $\pi(x)/x$, descending by powers of $\log x$.

Much effort has been exerted to learn what the smallest function is with which the right side of (3) can be replaced. In 1958, in independent work, I. M. Vinogradov [**22**] and N. M. Korobov [**13**] showed that the right side of (3) can be replaced by

$$c_4 x e^{-c_5(\log x)^b},$$

where $b$ is any constant less than $3/5$, and $c_4$ and $c_5$ are positive constants which may depend on $b$. It has been conjectured that in fact, the right side of (3) can be reduced to the quantity

$c_6\sqrt{x}\log x$, for some positive constant $c_6$. This improved version of (3) was shown by von Koch in 1901 [**12**] to be equivalent to the so-called Riemann Hypothesis. To state this hypothesis, we define the Riemann zeta-function by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for the real part of $s$ exceeding 1. This function has an analytic continuation to $\{s \text{ complex: } s \neq 1\}$, with a simple pole of residue 1 at $s = 1$. Its connection with primes can be seen in the beautiful formula due to Euler [**6**], which states that

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}.$$

Its proof is given on p. 231 of [**1**]. The Riemann Hypothesis is the conjecture that $\zeta(s)$ has no zeroes with real part exceeding $1/2$. It is supported by a very large amount of evidence [**15**], [**2**], [**3**].

Since there are roughly $x/\log x$ primes $p \leqslant x$, the average distance between consecutive primes not exceeding $x$ is $\log x$. If $p \geqslant 3$, then the next prime must be at least $p + 2$. It is conjectured that $p + 2$ is prime infinitely often. Pairs of primes of the form $p, p + 2$, like 3, 5 and 101, 103, are called twin primes, and the as yet unproven conjecture about the existence of infinitely many such pairs is called the Twin Prime Conjecture. J. R. Chen (see p. 4 and Chapter 11 of [**8**]) has shown that there are infinitely many $p$ for which $p + 2$ has at most two prime factors. If $p \geqslant 5$, then $p, p + 2, p + 4$ cannot all be prime, since 3 evenly divides one of them. However, it is conjectured that there are arbitrarily large $p$ for which $p, p + 2, p + 6$ are all prime. These conjectures can be generalized considerably (see Chapter 1 of [**8**]).

In the other direction, we can ask how large the gap between consecutive primes can be. There is conjectured to be a prime between any two squares of positive integers. Let $p^*$ denote the next prime after $p$. Then this conjecture would imply that $p^* < p + 4p^{1/2}$ for all $p$, and would follow if we could show that $p^* < p + 2p^{1/2} + 2$ for all $p$. In [**10**], D. R. Heath-Brown and H. Iwaniec showed that for any real number $\mathcal{O} > 11/20$, there is a constant $c_7$, possibly depending on $\mathcal{O}$, so that $p^* < p + c_7 p^{\mathcal{O}}$. In particular, there is a prime between any sufficiently large cubes of integers (this fact was already known to Ingham [**11**] in 1937).

The ancient Greeks defined a perfect number to be a positive integer $n$ which is the sum of all positive integers $d < n$ which evenly divide $n$. Thus, the first few perfect numbers are 6, 28, 496, 8128, 33550336. The ancient Greeks proved that every even perfect number has the form $2^{p-1}(2^p - 1)$, where $p$ and $2^p - 1$ are primes. It is still unknown whether any odd perfect numbers exist. The numbers $2^p - 1$ are called Mersenne numbers, after a seventeenth century French mathematician, Marin Mersenne, who listed values of $p$ for which he thought $2^p - 1$ to be prime. There are 29 values of $p$ now known for which $2^p - 1$ is prime, the largest of which is $p = 132{,}049$. The corresponding Mersenne prime—which is also the largest known prime—has slightly less than 40,000 digits, and the corresponding perfect number has slightly less than 80,000 digits. The second largest known value of $p$ which gives a Mersenne prime is 86,243. For a table of the others, see [**18**]. The values 132,049 and 86,243 were found by David Slowinski, with the aid of supercomputers made by Cray Research in Chippewa Falls, Wisconsin.

Finding large primes used to be of merely theoretical interest, until 1977, when Robert L. Rivest, Adi Shamir, and Leonard M. Adleman discovered an application to national security. In their system, two large primes, $p$ and $q$, are multiplied together. It follows from Euler's generalization of Fermat's Little Theorem that for any positive integer $a$, we have

$$a^{(p-1)(q-1)+1} \equiv a \pmod{pq}. \tag{4}$$

Hence, if we choose any two positive integers $b$ and $c$ with $bc = (p - 1)(q - 1) + 1$, then we can use the following code:

To send the information "$a$", for $0 \leqslant a < pq$, find the remainder of $a^b$ on division by

$pq$, and communicate that. To decode the information received, raise it to the $c$th power, and find the remainder on division by $pq$. That remainder is $a$.

For example, if $p = 13$, $q = 19$, then

$$(p-1)(q-1)+1 = 217 = 7 \cdot 31,$$

so that we can take $b = 31$, $c = 7$, $pq = 247$. To send "2", we observe that $2^{31} \equiv 193 \pmod{247}$, and communicate "193". The receiver then notes that $193^7 \equiv 2 \pmod{247}$, to decode the message. The elegant thing about this type of code is that even if the enemy learns how to encode messages, it is much easier to find two primes with—say—200 digits each, and multiply them together, than to factor the resulting 400 digit number. It will shortly be possible to determine if a 200 digit number is prime in 8 minutes [**19**]. To *factor* even a *100* digit number can currently take decades. For a fuller account, we refer the reader to [**18**].

Although the book on which this article is based was not written by a pure mathematician, it contains a great amount of historical information about the lore of prime numbers. The book contains few, if any, proofs. Except for a few sections, all that a student needs to read it is a good understanding of the first year of high school algebra, although it should also be enjoyable reading material for the undergraduate student who has had a semester of elementary number theory.

## References


**1.** T. M. Apostol, Introduction to Analytic Number Theory, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1976, xii + 338 pp.

**2.** R. P. Brent, J. van de Lune, H. J. J. te Riele, and D. T. Winter, The first 200,000,001 zeroes of Riemann's zeta function, Math. Comp., 39 (1982) no. 160, 681–688.

**3.** ______, The first 200,000,001 zeroes of Riemann's zeta function, Computational Methods in Number Theory, part II, 389–403, Math. Center Tracts, 155, Math. Centrum. Amsterdam, 1982.

**4.** R. C. Buck, Prime-representing functions, this MONTHLY, 53 (1946) 265.

**5.** M. Davis, Hilbert's tenth problem is unsolvable, this MONTHLY, 80 (1973) 233–269.

**6.** L. Euler, Variae observationes circa series infinitas, Commentarii Academiae Scientiarum Imperialis Petropolitanae, 9 (1737) 160–188. [Opera Omnia (1), 14: 216–244.]

**7.** J. Hadamard, Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques, Bull. Soc. Math. France, 24 (1896) 199–220.

**8.** H. Halberstam and H. -E. Richert, Sieve Methods, L.M.S. monographs, no. 4, Academic Press, New York, 1974, xiv + 364 pp.

**9.** G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960, xvi + 421 pp.

**10.** D. R. Heath-Brown and H. Iwaniec, On the difference between consecutive primes, Invent. Math., 55 (1979) no. 1, 49–69.

**11.** A. E. Ingham, On the difference between consecutive primes, Quart. J. Math. Oxford Ser., 8 (1937) 255–266.

**12.** H. von Koch, Sur la distribution des nombres premiers, Congrès Intern. des Math., Paris, 1900, 195–198 (1902).

**13.** N. M. Korobov, Weyl's estimates of sums and the distribution of primes, Dokl. Akad. Nauk SSSR, 123 (1958) 28–31. (Russian.)

**14.** W. J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, MA, 1977, vii + 280 pp.

**15.** N. Levinson, More than one third of zeroes of Riemann's zeta-function are on $\sigma = 1/2$, Adv. in Math., 13 (1974) 383–436.

**16.** G. P. Loweke, The Lore of Prime Numbers, Vantage Press, New York, 1982, vii + 259 pp.

**17.** W. H. Mills, A prime-representing function, Bull. Amer. Math. Soc., 53 (1947) no. 6, 604.

**18.** C. Pomerance, The search for prime numbers, Scientific American, 247 (Dec. 1982) no. 6, 136–147.

**19.** ______, Personal communication, Western Number Theory Meeting (Monterey, California, December 1983).

**20.** Charles de la Vallée Poussin, Recherches analytiques sur la théorie des nombres premiers, Ann. Soc. Sci. Bruxelles, $20_2$ (1896) 183–256.

**21.** ______, Sur la fonction $\zeta(s)$ de Riemann et le nombre des nombres premiers inférieurs à une limite donnée, Mém. Couronnés Acad. Roy. Belgique, 59 (1899) 1–74.

**22.** I. M. Vinogradov, A new estimate of the function $\zeta(1+it)$, Izv. Akad. Nauk SSSR. Ser. Mat., 22 (1958) 161–164. (Russian.)

*Problem-Solving through Problems*. By Loren C. Larson. Problem Books in Mathematics. Springer-Verlag, New York, 1983. xi + 332 pp. $34.00.

HARLEY FLANDERS
*Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431*

Springer-Verlag, the most important mathematics publishing house of the twentieth century, the house that brought us the incomparable Pólya and Szegő, has now, in its new series *Problem Books in Mathematics*, published the volume under review. The stated purpose of the book is "to isolate and draw attention to the most important problem-solving techniques... undergraduate mathematics... interesting examples and problems not easily found in other sources... material is especially appropriate for students preparing for mathematical competitions."

When I received the volume, I looked forward to finding some interesting new problems, some excellent solutions, some really useful text material for students. I shall try to share my disappointments.

Chapter 1 has twelve *heuristics*: *Search for a Pattern*, *Exploit Symmetry*, *Choose Effective Notation*, etc., each illustrated by several, often inappropriate, worked examples and some problems. (I can't understand the lack of an answer section—what good are the roughly 470 exercises with solutions scattered in the literature or not even referenced?)

Under *Search for a Pattern*, the author wisely gives four examples in which to search for a pattern means to examine low cases. Then follows his fifth example (1.1.5, page 7): *Let $S$ be a set and $*$ be* (*a*) *binary operation on $S$ satisfying the two laws $x * x = x$ for all $x$ in $S$, $(x * y) * z = (y * z) * x$ for all $x, y, z$ in $S$. Show that $x * y = y * x$ for all $x, y$ in $S$.* In spite of what the author says, solving this problem really seems a different matter, and the author gives essentially no insight into how to solve such problems. (Using a lower case $x$ next to $*$ is a model for confusing notation.)

I think that in a book like this, the author usually ought to tell his readers what are hard problems and what are not so hard problems. In the heuristic *Consider Extreme Cases*, the very first example is Problem 1.11.1 (page 50): *Given a finite number of points in the plane*, *not all collinear*, *prove there is a straight line which passes through exactly two of them*. That this yields to minimizing the distance between a point $P$ of the set and a line through two other points of the set, but not through $P$, does not detract from the fact that it is a *very* hard problem that took 40 years from Sylvester's 1893 conjecture to Gallai's solution. The author could tell us this story.

Nowhere in the heuristics chapter, or anywhere else, is there a mention of George Pólya. This is unforgiveable.

The chapters after the first take up specific topics and apply the material to problem solving. For instance there is a chapter on inequalities, a chapter on summation of series, a chapter on geometry, etc. I find time and again that the examples are misplaced and have better solutions by other techniques, discussed in this very book. What I find particularly annoying are long, cloddish, computational solutions. Frequently the author gives the distinct impression that grinding out computation is the real heart and difficulty of a solution, whereas the essential idea just drops effortlessly out of "heuristics" and is passed off with "Notice that...". Surely every single solution in a book whose purpose is to teach problem solving should emphasize the key ideas and never snow the reader with calculations. No solution should be published if it is clumsy or there remains any hope of replacing computation by concepts. (It is possible that no mathematician besides the author checked the manuscript before publication—no one's help is acknowledged in the Preface. The large number of misprints suggests this and a careless proofreading job.)

I happened to be working with some students on preparation for the Putnam Competition when Larson's book appeared. I used a number of his solutions as models of how not to solve problems. The book is a rich vein for such models; perhaps this is the most appropriate use of the text. Some examples follow; they are not exceptional, but typical, cases. Consider Problem 3.5.5, page 117:

*Let $G_n = x^n \sin nA + y^n \sin nB + z^n \sin nC$, where $x, y, z, A, B, C$ are real and $A + B + C$ is an integral multiple of $\pi$. Prove that if $G_1 = G_2 = 0$, then $G_n = 0$ for all positive integral $n$.* (Author's source: 1980 USA Olympiad.)

The author's solution starts with the reasonable observation that $G_n$ is the imaginary part of $x^n e^{inA} + y^n e^{inB} + z^n e^{inC}$, but then gives a horrendous, brute-force, unmotivated computation that in no way shows where the result is coming from. Yet, if we set

$$\alpha = xe^{iA}, \qquad \beta = ye^{iB}, \qquad \gamma = ze^{iC},$$

it is natural to consider

$$f(T) = (T-\alpha)(T-\beta)(T-\gamma) = T^3 - \sigma_1 T^2 + \sigma_2 T - \sigma_3$$

and the sums of powers $s_n = \alpha^n + \beta^n + \gamma^n$. Now we are on well-trodden ground, and in case we have forgotten Newton's relations

$$s_{n+3} - \sigma_1 s_{n+2} + \sigma_2 s_{n+1} - \sigma_3 s_n = 0,$$

then multiplying $f(T)$ by $T^n$ and summing on $T = \alpha, \beta, \gamma$ will jog our memories. The hypotheses easily imply with almost no computation that $\sigma_1 = s_1$ is real, $\sigma_3$ is real, $s_2$ is real, so $2\sigma_2 = s_1^2 - s_2$ is real. So is $s_0 = 3$, and then by induction so is $s_n$.

Even worse is the author's solution to Problem 5.3.3, page 172:

*Express*

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} 1/[m^2 n + mn^2 + 2mn]$$

*as a rational number.* (Author's source: 1978 Putnam competition, B2. With all due respect to the Putnam, "Evaluate" without "rational number" would have been a better statement.)

The author's solution is a full page orgy of numerical computation, starting with a partial fraction decomposition of the denominator $mn(m+n+2)$. The final answer seems like a miracle. Aside from the obvious ugliness of an unbroken string of complicated equalities stretching over 17 lines, with at least one misleading misprint, all of which could have been organized into something far simpler and intelligible, how on earth would the author's method have applied to denominator $mn(m+n+7)$ or $mnp(m+n+p+3)$? (Heuristic 12: *Generalize!*)

Now if we, very naturally, set

$$f(x) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x^{m+n+k+1}/[mn(m+n+k+1)],$$

then one differentiation clears away most of the foliage, and we have

$$\begin{aligned} S_k &= \sum_1^{\infty} \sum_1^{\infty} \frac{1}{mn(m+n+k+1)} = \int_0^1 x^k [\ln(1-x)]^2\,dx \\ &= \int_0^1 (1-x)^k (\ln x)^2\,dx. \end{aligned}$$

It is a routine binomial expansion and integration from here to

$$S_k = 2\sum_{j=0}^{k} \frac{(-1)^j}{(j+1)^3} \binom{k}{j}.$$

Now let's look at Problem 6.6.4, page 219:

*Let $f$ be differentiable with $f$ continuous on $[a,b]$. Show that (if) there is a number $c$ in $(a,b]$ such that $f'(c) = 0$, then we can find a number $\xi$ in $(a,b)$ such that $f'(\xi) = [f(\xi) - f(a)]/(b-a)$.*

(Author's source: S. Penner, Math. Mag., 49, p. 150.)

The solution fills two full pages, is repetitious and I think misses the point. First of all, the denominator $b - a$ can be anything. Second of all, $\xi$ can be found in $(a, c]$. Thus (Heuristic 4: *Modify the Problem*) the problem should be:

*Let $f(x)$ be continuous in $[a, b]$, differentiable on $(a, b]$, with $f'(b) = 0$, and let $K > 0$. Prove there is an $x$ in $(a, b)$ such that $f'(x) = K[f(x) - f(a)]$.*

Clearly there is a simplification of notation (Heuristic 5) and no loss of generality in taking $a = 0$ and $f(a) = 0$. Set $g(x) = e^{-Kx}f(x)$, a device the author mentioned a few pages earlier. If $f'(x) > Kf(x)$ on $(a, b)$, then $g'(x) > 0$ so $g(b) > g(a) = 0$, hence $f(b) > 0$. By the Darboux property, $f'(b) \geqslant Kf(b) > 0$, which is impossible. Similarly, $f'(x) < Kf(x)$ on $(a, b)$ is impossible, so the conclusion follows.

In the sections on geometry, the author outdoes himself with clumsy notation and computational instead of conceptual solutions. Imagine using $|\overrightarrow{PQ}|$ for the length of $\overrightarrow{PQ}$ and next to it $|\overrightarrow{PQ}$ for the vector obtained by rotating $\overrightarrow{PQ}$ about $P$ through $90°$. See p. 306 where such awkward notation almost forces misprints. Let's look at a nice little problem from the 1977 USA Olympiad, Problem 8.3.8, page 308:

*Prove that if the opposite sides of a skew quadrilateral have the same lengths, then the line joining the midpoints of the two diagonals is perpendicular to these diagonals.*

The author's start is good: Let $A, B, C, D$ denote the vertices of the quadrilateral (in order), and let $P$ and $Q$ be the midpoints of $AC$ and $BD$ respectively. Now I'll replace 1 1/2 pages of computation by a conceptual solution (Heuristic 6: *Exploit Symmetry*): Let $\rho$ denote the affine reflection in the plane $BPD$ that transposes $A$ and $C$. Let $\sigma$ denote the affine reflection in the plane $AQC$ that transposes $B$ and $D$. Set $\lambda = \sigma\rho$. Then

$$\lambda\colon A \to C, \quad C \to A, \quad B \to D, \quad D \to B, \quad P \to P, \quad Q \to Q.$$

By hypothesis, $\overline{AB} = \overline{CD}$ and $\overline{BC} = \overline{DA}$, so it follows that $\lambda$ preserves the lengths of the six sides of the tetrahedron $ABCD$. Therefore $\lambda$ is a rigid motion, clearly not a reflection. Since $\lambda$ preserves $P$ and $Q$, it is a rotation about $PQ$. Since $\lambda$ transposes $B$ and $D$, $BD$ is perpendicular to $PQ$. Similarly so is $AC$.

I conclude that the author has failed to meet his goals. Some of his problems, particularly the many taken from Putnam competitions and Olympiads are excellent, but not at all hard to find in the literature. As for problems from the periodical literature and other sources: some are good and some are bad; it is not an outstanding collection. As for the solutions: The long ones can usually be shortened; the short ones can be incomplete, even incorrect. Consider Problem 2.5.4, page 76:

*Prove that any positive rational number can be expressed as a finite sum of distinct terms of the harmonic series.* (Author's source: 1954 Putnam, B6.)

This famous theorem on Egyptian fractions is a tricky proposition to prove, and the author could at least have consulted the Putnam Competition book, p. 398, if not one of the many number theory books that supply a proof. Not at all; he polishes it off in a few lines: "Let $m/n$ be any positive rational. Then $m/n = 1/n + 1/n + \cdots + 1/n$ is a sum of harmonic terms with $n - 1$ (sic) duplications. Recursively expand all duplicates by the identity $1/n = 1/(n + 1) + 1/n(n + 1)$ until all terms are distinct." Aside from the misprint, this is no proof at all, and I see no way to make it into a proof. The very last problem on a Putnam competition is seldom trivial at all, let alone *that* trivial. New heuristic: *Check Your Work.*

(The author is grateful to Murray S. Klamkin and Stephen C. Locke for helpful suggestions.)

# LETTERS TO THE EDITOR

*Material for this department should be prepared exactly the same way as submitted manuscripts (see the inside front cover) and sent to Professor P. R. Halmos, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.*

Editor:

While I found the approach taken by Professor Meyer Dwass in his note on the convolution of Cauchy distributions [**1**] both novel and interesting, I am not convinced that it is simpler algebraically than the direct method to which he refers; namely the partial fraction decomposition of the integrand in that convolution. I, like Professor Dwass, could not find the details worked out anywhere; and so perhaps this would be the time and place to present them.

The convolution integral to be evaluated is

$$h(u) = \int_{-\infty}^{\infty} f(x,a) f(u-x,b)\,dx,$$
$$= \frac{ab}{\pi^2}\int_{-\infty}^{\infty} \frac{1}{[a^2+x^2]\left[b^2+(u-x)^2\right]}\,dx.$$

The partial fraction decomposition of the integrand is

$$\frac{1}{[a^2+x^2]\left[b^2+(u-x)^2\right]} = \frac{2Ax+B}{a^2+x^2} + \frac{2C(u-x)+D}{b^2+(u-x)^2},$$

with

$$A = C = u/\Delta, \quad B = (u^2-a^2+b^2)/\Delta, \quad \text{and} \quad D = (u^2+a^2-b^2)/\Delta,$$

where the denominator in each coefficient is

$$\Delta = \left[(a+b)^2+u^2\right]\left[(a-b)^2+u^2\right].$$

Integration now yields

$$h(u) = \frac{1}{\pi^2\Delta}\left[abu\ln\left(\frac{a^2+x^2}{b^2+(u-x)^2}\right) + b(u^2-a^2+b^2)\tan^{-1}\frac{x}{a}\right.$$
$$\left. -a(u^2+a^2-b^2)\tan^{-1}\frac{u-x}{b}\right]_{-\infty}^{\infty}$$
$$= \frac{1}{\pi\Delta}\left[b(u^2-a^2+b^2)+a(u^2+a^2-b^2)\right]$$
$$= \frac{a+b}{\pi\left[(a+b)^2+u^2\right]} = f(u,a+b), \quad \text{as desired.}$$


**Reference**

1. Meyer Dwass, On the convolution of Cauchy distributions, this MONTHLY, 92 (1985) 55–57.



Roger B. Nelsen
Department of Mathematics
Lewis and Clark College
Portland, OR 97219

**156.** **MISCELLANEA**

Nothing is more repellent to normal human beings than the clinical succession of definitions, axioms, and theorems generated by the labours of pure mathematicians. The logical rigour achieved by such investigations is of the highest value, but can seldom come before we have grasped the idea in itself. Geometry existed before Euclid, and analysis before Cauchy; the significance of an irreducible representation of a group can be understood without a proof of Schur's Lemma in all its generality.

—J. M. Ziman, *Elements of Advanced Quantum Theory* (Cambridge, 1969), p. vii.

---

**157.** **MISCELLANEA**

I maintain that we must be careful of a problem's first contact with our minds. We should be careful of the first words a question utters in our minds. A new question arising in us is in a state of infancy; it stammers; it finds only strange terms, loaded with adventitious values and associations; it is forced to borrow these. But it thereby insensibly deflects our true need. Without realizing it we desert our original problem, and in the end we shall come to believe that we have chosen an opinion wholly our own, forgetting that our choice was exercised only on a mass of opinions that are the more or less blind work of other men and of chance.

—Paul Valéry, "Poetry and Abstract Thought," *Paul Valéry, an Anthology*, Princeton University Press, 1977, p. 138.

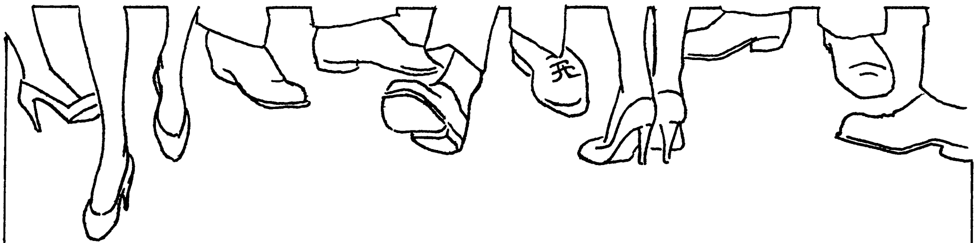
---

**158.** **MISCELLANEA**

**Clerihew Corner**

George Boole
was nobody's fool:
but never forget—
his mathematical legacy is the empty set.

Robin Harte
Department of Mathematics
University College
Cork, Ireland



U S Postal Service

**STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION**

*Required by 39 U S C 3685)*

1A TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY

1B PUBLICATION NO: A M M Y A E

2 DATE OF FILING: Sept. 16, 1985

3 FREQUENCY OF ISSUE: monthly except bi-monthly June/June and Aug./Sept.

3A NO OF ISSUES PUBLISHED ANNUALLY: Ten

3B ANNUAL SUBSCRIPTION PRICE: Library $55 Indiv.Member $22

4 COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street City County State and ZIP+4 Code) (Not printers)*

1529 Eighteenth Street, N.W., Washington, D.C. 20036

5 COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OF GENERAL BUSINESS OFFICES OF THE PUBLISHER *(Not printer)*

1529 Eighteenth St., N.W., Washington, D.C. 20036

6 FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR AND MANAGING EDITOR *(This item MUST NOT be blank)*

PUBLISHER *(Name and Complete Mailing Address)*

The Mathematical Association of America, 1529 Eighteenth St., N.W., Washington, D.C. 20036

EDITOR *(Name and Complete Mailing Address)*

P.R. Halmos, Univ. of Santa Clara, Dept. of Mathematics, Santa Clara, CA 95053

MANAGING EDITOR *(Name and Complete Mailing Address)*

Dr. Raoul Hailpern, 63 Garden Court, Amherst, N.Y. 14226

7

| FULL NAME | COMPLETE MAILING ADDRESS |
|---|---|
| The Mathematical Association of America, Inc. | 1529 Eighteenth St., NW Washington, D.C. 20036 |

8 KNOWN BONDHOLDERS MORTGAGEES AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none so state)*

| FULL NAME | COMPLETE MAILING ADDRESS |
|---|---|
| None | |

9 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 423 12 DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS

(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS

*(If changed publisher must submit explanation of change with this statement )*

| 10 EXTENT AND NATURE OF CIRCULATION *(See instructions on reverse side)* | AVERAGE NO COPIES EACH ISSUE DURING PRECEDING 12 MONTHS | ACTUAL NO COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE |
|---|---|---|
| A TOTAL NO COPIES *(Net Press Run)* | 20,425 | 19,175 |
| B PAID AND/OR REQUESTED CIRCULATION 1 Sales through dealers and carriers street vendors and counter sales | 100 | 100 |
| 2 Mail Subscription *(Paid and/or requested)* | 17,325 | 17,076 |
| C TOTAL PAID AND/OR REQUESTED CIRCULATION *(Sum of 10B1 and 10B2)* | 17,425 | 17,176 |
| D FREE DISTRIBUTION BY MAIL CARRIER OR OTHER MEANS SAMPLES COMPLIMENTARY AND OTHER FREE COPIES | 65 | 70 |
| E TOTAL DISTRIBUTION *(Sum of C and D)* | 17,490 | 17,246 |
| F COPIES NOT DISTRIBUTED 1 Office use, left over, unaccounted spoiled after printing | 2,935 | 1,929 |
| 2 Return from News Agents | 0 | 0 |
| G TOTAL *(Sum of E F1 and 2–should equal net press run shown in A)* | 20,425 | 19,175 |

11 I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR PUBLISHER BUSINESS MANAGER OR OWNER

PS Form 3526, July 1984 *(See instruction on reverse)*

# INDEX TO VOLUME 92, 1985
# THE AMERICAN MATHEMATICAL MONTHLY



## AUTHOR INDEX

## SUBJECT INDEX

This index uses the current version of AMS 1980 Mathematics Subject Classification.

**53-XX DIFFERENTIAL GEOMETRY**



**54-XX GENERAL TOPOLOGY**



**57-XX MANIFOLDS AND CELL COMPLEXES**



**58-XX GLOBAL ANALYSIS, ANALYSIS ON MANIFOLDS**



**60-XX PROBABILITY THEORY AND STOCHASTIC PROCESSES**



**62-XX STATISTICS**



**65-XX NUMERICAL ANALYSIS**



**68-XX COMPUTER SCIENCE (INCLUDING AUTOMATA)**



**70-XX MECHANICS OF PARTICLES AND SYSTEMS**



**90-XX ECONOMICS, OPERATIONS RESEARCH, PROGRAMMING, GAMES**

## PROBLEMS AND SOLUTIONS

### PROBLEMS PROPOSED



### PROBLEMS SOLVED

## SOLUTIONS

*Numbers in boldface refer to problems; those in lightface to pages.*



## REVIEWS

Names of authors are in ordinary type; those of reviewers in capitals.

## LETTERS TO THE EDITOR



## PHOTOS



## MISCELLANEA



## INDEX OF SPECIAL RECURRING TOPICS



## REBUSES

# THE AMERICAN MATHEMATICAL MONTHLY

Volume 92, Number 10 December 1985


## Contents

(ISSN 0002-9890)

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1980 and later. (This is necessary for indexing purposes.) Send all main articles, letters to the editor, and editorial correspondence to the editor. Contributions to the other four special departments (Notes, Teaching of Mathematics, Unsolved Problems, and Problems and Solutions) should be sent to the appropriate one of the associate editors, whose addresses are given below.

---



Notes. Combinatorics and probability: SABRA S. ANDERSON, Department of Mathematics, University of Minnesota, Duluth, MN 55812. Analysis and linear algebra: SHELDON AXLER, Department of Mathematics, Michigan State University, East Lansing, MI 48824. Topology, geometry, and abstract algebra: J. ARTHUR SEEBACH, JR., Department of Mathematics, St. Olaf College, Northfield, MN 55057. Authors of fields not covered by these rubrics may send their notes to any one of the three editors.

Teaching of Mathematics: ROBERT F. WARDROP, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI 48859.

Unsolved Problems: RICHARD GUY, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Proposed problems and solutions, both Elementary and Advanced: D. H. MUGLER, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053.

Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Back Issues: P. and H. BLISS Co., Middletown, CT 06457.

Microfilm Editions: University Microfilms International, Serials Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

Editorial Assistants: Fanny Hailpern, Mary P. Jackson.

---


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT, ten times a year: January, February, March, April, May, June-July, August-September, October, November and December.

The annual subscription price for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is $22 included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $22. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $55 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

PRINTED IN THE UNITED STATES OF AMERICA

# LOCAL SYMMETRY OF PLANE CURVES

P. J. GIBLIN AND THE LATE S. A. BRASSETT
*Department of Pure Mathematics, The University of Liverpool, P.O. Box 147, Liverpool L69 3BX, England*

**1.** How can we perceive, recognize, describe or represent the shape of an object? These questions are of great interest to workers in theoretical biology and in the perception of patterns —whether by people or by computers. Organisms have simple shapes which are not usually equilateral triangles, squares or other basic mathematical shapes, so new concepts are needed to capture their simplicity. Organisms move and grow: what is preserved as their shapes change in these ways? Shapes that we all immediately recognize as the same (such as all the instances of the letter m in this journal) are in fact subtly different: what do they have in common?

In order to tackle questions such as these, a number of ways have been suggested for reducing the huge amount of information carried by a shape down to a "skeleton" of crucial, possibly discrete, information which can be more readily assimilated. For example, H. Blum [**1**] suggested that a planar shape could be studied by fitting circular disks inside it and noting the locus of their centres, which he called the "sym-ax" or symmetric axis transform (Fig. 1, left). A wiggling worm can be described, perhaps, by keeping the radii fixed and flexing the sym-ax; a growing worm by changing the radii but leaving the sym-ax unaltered.

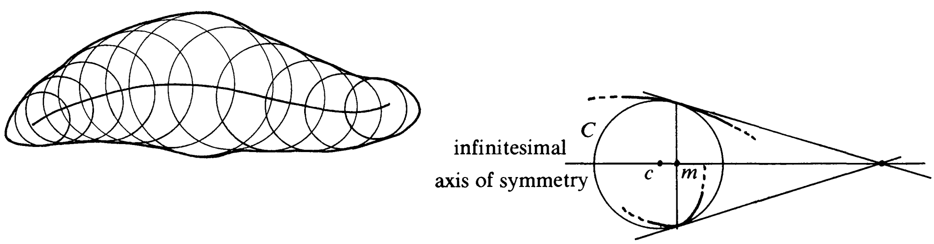



FIG. 1

The mathematical questions raised by such an imaginative idea are very interesting. Suppose we take a planar shape, bounded, let us say, by a simple closed curve. What forms will the locus of centres take?

Whenever a circle is bitangent to a curve, that is tangent to the curve at two different places (Fig. 1, right), there is an "infinitesimal axis of symmetry" which bisects the angle between the tangents to the curve (or the circle) at the two places. The centre $c$ of the circle lies on Blum's sym-ax, at any rate if the circle lies inside the curve close to the points of contact. For mathematical purposes we prefer to drop the latter requirement, that is we allow all bitangent circles including circles such as those in Fig. 5. We also change the name to *symmetry set*. Thus:

*The symmetry set of a plane curve is the locus of centres c of circles bitangent to the curve.*

Of course in special cases the circle could be tangent to the curve at three (or more) different places. Taking three places in pairs, such a point would lie on three "branches" of the symmetry set (and for four places, six branches).

---

Peter Giblin joined the faculty at Liverpool University in 1967 and has been there ever since, apart from a year as Visiting Professor at the University of North Carolina at Chapel Hill. His special interests are in the geometrical applications of singularity theory, many of which can be effectively illustrated, and even aided, by the use of computer graphics.

Stephen Brassett was a final year undergraduate at Liverpool during the time that the work described in this article was done. He wrote a project, under Giblin's supervision, in which he used computer graphics to study symmetry sets and midloci. Besides computer programming, Stephen had an intense interest in butterflies. From the time of his graduation in 1984 until his tragically early death in January 1985 he worked in industry in the field of computer aided design.

Because the symmetry set, or the sym-ax, suffers from various technical shortcomings (for example it reflects poorly the perceived symmetry of a rectangle), M. Brady in [2] suggested a simple variant: instead of the centre of the circle we take the mid-point $m$ of the chord of contact. (The reader will find many more references to follow up in [1] and [2].) He called the resulting locus the "smoothed local symmetry", but we shall adopt a more directly descriptive term.

*The midpoint locus (midlocus for short) of a plane curve is the locus of midpoints $m$ of chords of contact of bitangent circles.*

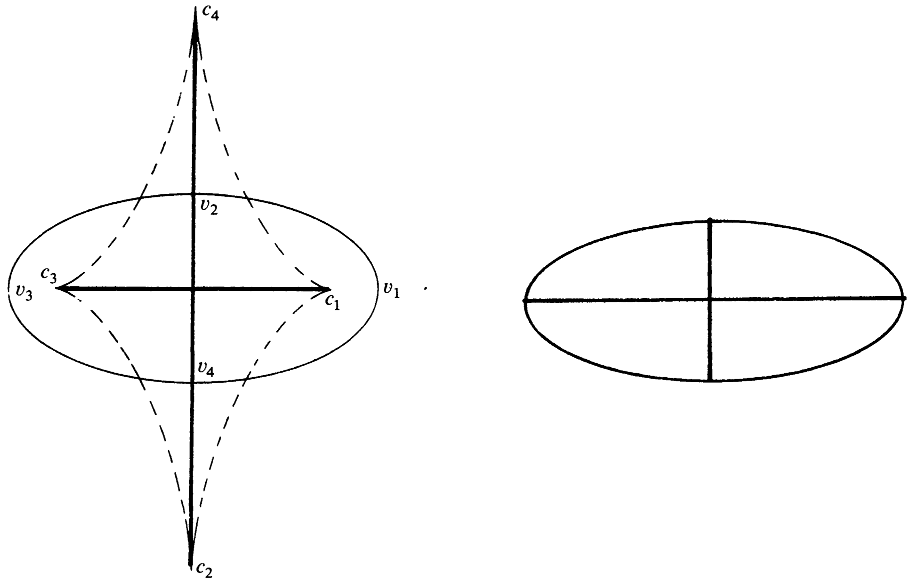



FIG. 2. Symmetry set and midlocus for an ellipse.

Let us consider a very simple example, when the curve is an ellipse (Fig. 2), say given by $\frac{1}{4}x^2 + y^2 = 1$. Evidently—by symmetry!—we need only consider circles centred on one or other coordinate axis; let us take the $x$-axis for which the contact is internal to the ellipse. The circle $(x - a)^2 + y^2 = r^2$ has slope $-(x - a)/y$ at $(x, y)$ and the ellipse has slope $-x/4y$ at $(x, y)$, so if they meet at a point $(x, y)$, where the tangents coincide, then $x = 4a/3$. From the equation of the ellipse we find $y^2 = 1 - 4a^2/9$, so that $|a| \leqslant 3/2$, and the circle touches at $(4a/3, \pm(1 - 4a^2/9)^{1/2})$. Thus the portion of the $x$-axis between $-3/2$ and $3/2$ is in the symmetry set. Since the midpoint of the chord of contact is at $(4a/3, 0)$, the midlocus contains the interval from $-2$ to $2$ on the $x$-axis, that is precisely, and not surprisingly, the major axis of the ellipse. The full symmetry set and midlocus are the solid lines in Fig. 2. The crossover point on the symmetry set is the centre of two concentric circles each bitangent to the ellipse. Notice also that, as $a \to 3/2$, the points of contact of circle and ellipse tend to coincidence at $(2, 0)$: the circle centre $(3/2, 0)$ has *all four* points of intersection with the ellipse coincident at $(2, 0)$. This *four-point contact* results in an abrupt end to the symmetry set and midlocus. Thus the four points $e_i$ in Fig. 2 are strictly in the closure of the symmetry set and the corresponding points $v_i$, the *vertices* of the ellipse, in the closure of the midlocus, as these are defined above.

The dashed curve in Fig. 2 is the *evolute* of the ellipse, that is the locus of centres of circles for which three of their intersections with the ellipse have coincided at a single point. These are circles with *three-point contact*, or osculating circles, or circles of curvature. Note that the evolute has cusps at the points $c_i$ corresponding to vertices $v_i$.

Fig. 3 shows a more complex example, but with the same "cross" shape to the symmetry set and midlocus. Again the $v_i$ are vertices, where a certain circle has four-point contact with the

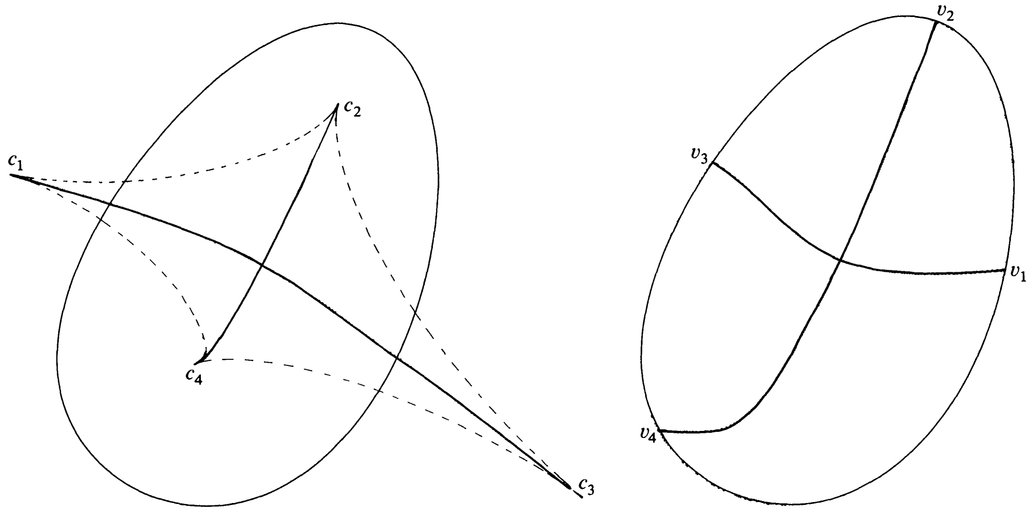



FIG. 3. Symmetry set and midlocus for a cubic oval $(x - 0.9y)^2 = y - y^3$.

curve; the centre of this circle is at $c_i$ on the evolute.

Because we have to consider *pairs* of points of the curve, the constructions of the symmetry set and midlocus are altogether more complicated than that of, say, the evolute. They give rise not only to interesting mathematical problems but also to computational problems, and we say a little about the latter in §2.

Suppose we are given a smooth simple closed plane curve and a point $p$ on it. Is there always a circle or straight line tangent to the curve at $p$ and at another point $p' \neq p$? In fact the answer is yes, and we invite the reader to find a simple proof. This shows that to every point of such a curve there corresponds either a point of the symmetry set or a limit point "at infinity" (corresponding to a bitangent *line*). The point is always *finite* for the midlocus, for we can take the midpoint of the chord of contact in the case of a bitangent line (compare Fig. 18, right).

In **[5]** an elaborate mathematical tool is used to study the symmetry set for curves in the plane and also surfaces in 3-space (where *circles* become *spheres*). One of the results is the following (see also **[3]** p. 137). Consider a point $x$ in the plane, on the symmetry set of a generic plane curve (i.e., practically any curve). Then, after a diffeomorphism of a neighbourhood of $x$ in the plane, the symmetry set looks like one of the five pictures of Fig. 4. Note that the triple crossing, corresponding to a circle tangent at three places to the curve, cannot be removed by a small perturbation of the curve. Note also that the last picture shows a smooth piece of the symmetry set coming to an end (as with the ellipse and cubic oval above). For another theoretical investigation of the related concept of *central set* see **[11]**.

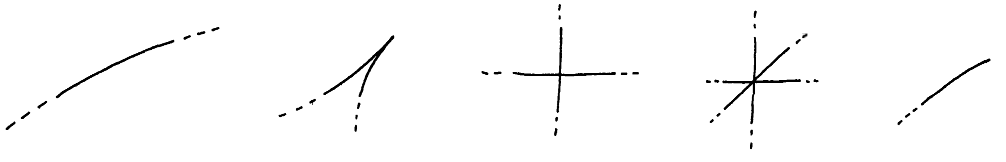


FIG. 4. The five local pictures for the symmetry set of a generic plane curve.

Our purpose here is much less technical. We apply some techniques from calculus, notably the implicit function theorem, to describe the symmetry set and midlocus locally in a variety of situations, illustrating by means of computer drawn examples. The midlocus is not covered at all in **[5]** and also we consider here a number of "non-generic" examples of the symmetry set. For

example if there is a "biosculating circle", that is, one having three-point contact with the curve at two separate places, then both midlocus and symmetry set can have an *isolated* point (see (3.1), (6.3) below). This occurs naturally when the curve is undergoing a deformation and we give an example of this in §7, where one part of the symmetry set shrinks to a point and vanishes (Fig. 26).

However, the midlocus has the pleasant property that it is very often smooth (free from cusps), though it can have endpoints and crossings as in Figs. 2 and 3. See (3.1).

One topic not covered here but treated in detail in [5] is the *recovery* of the curve from its symmetry set, by means of an envelope of circles. Another interesting idea is to look directly at the infinitesimal axes of symmetry, perhaps as a locus in the dual plane whose points are the lines of $\mathbb{R}^2$. For example, one can show that this dual locus acquires a cusp precisely when the points of contact of bitangent circles are points of *equal curvature* on the original curve—a pleasantly symmetrical condition. See (4.2) (1) below.

The plane of the paper is as follows. In §2 we review some results on the contact of circles with curves and set up various functions and maps with which we shall examine the symmetry set and midlocus. In §3 we deduce that the midlocus, and in §4 that the symmetry set, are smooth curves in suitable circumstances. In §5 we introduce the fold map from catastrophe theory and apply this to the cases where the midlocus has cusps. In §6 we look similarly at the singularities of the symmetry set. Finally in §7 we look briefly at families of curves and their evolving symmetry sets and midloci.

The computer pictures were all done using an IBM4341 and CalComp incremental plotter at Liverpool University.

**2. Contact of circles with curves.** A general reference for this section is [3]. Let $\gamma : I \to \mathbb{R}^2$ be a parametrization of a smooth plane curve, that is, $\gamma(t) = (x(t), y(t))$ where $x$ and $y$ are smooth functions and $x'(t), y'(t)$ are never zero for the same $t$. Here $I$ is an interval of real numbers, or possibly a circle for closed curves, but below we stick to intervals. The circle, with centre $u_0$ in the plane $\mathbb{R}^2$, and passing through the point $\gamma(t_0)$, has equation

$$\|x - u_0\|^2 - \|\gamma(t_0) - u_0\|^2 = 0,$$

double bars denoting length and $x$ here denoting the vector previously called $(x, y)$. This circle meets the curve $\gamma$ at points with parameter $t$ where

$$F(t, u_0) \equiv \|\gamma(t) - u_0\|^2 - \|\gamma(t_0) - u_0\|^2 = 0.$$

Of course $F(t_0, u_0) = 0$ by construction. If also $(\partial^i F/\partial t^i)(t_0, u_0) = 0$ for $i = 1, \ldots, k$, but $\neq 0$ for $i = k + 1$, then we say that the circle has $(k + 1)$-point contact with the curve at $\gamma(t_0)$, or has $A_k$ contact with the curve there. (The function $t \to F(t, u_0)$ has at $t_0$ an "$A_k$ singularity" in standard terminology; compare [3], p. 43.) Dropping the "$\neq 0$" condition above we say there is $A_{\geqslant k}$ contact. The following will be assumed in this paper—see for example [3], Chapter 2.

(2.1) *Write $T, N, \kappa$ for the unit tangent, unit normal and curvature of $\gamma$. Then, between the curve $\gamma$ and the circle with centre $u_0$ passing through $\gamma(t_0)$ there is:*

$A_{\geqslant 1}$ *contact if and only if $u_0$ lies on the normal to $\gamma$ at $t_0$ (i.e., $u_0 = \gamma(t_0) + \lambda N(t_0)$, for some real number $\lambda$).*

$A_{\geqslant 2}$ *contact if and only if $u_0$ is the centre of curvature of $\gamma$ at $t_0$ (i.e., $u_0 = \gamma(t_0) + N(t_0)/\kappa(t_0)$).*

$A_{\geqslant 3}$ *contact if and only if there is $A_{\geqslant 2}$ contact and $\gamma$ has a vertex at $t_0$ (i.e., $\kappa'(t_0) = 0$).*

$A_{\geqslant 4}$ *contact if and only if there is $A_{\geqslant 3}$ contact and $\gamma$ has a higher vertex at $t_0$ (i.e., $\kappa''(t_0) = 0$ too).*

Now the *evolute* of $\gamma$ is the locus of centres of curvature, that is the locus of points $u_0$ which are $A_{\geqslant 2}$ points for $\gamma$, meaning points for which the above circle has $A_{\geqslant 2}$ contact with $\gamma$ for *some*

$t_0$. These circles with $A_{\geqslant 2}$ contact somewhere are the *osculating circles* or *circles of curvature* of $\gamma$.

Both the symmetry set and the midlocus are concerned with circles not merely touching $\gamma$ (i.e., being tangent to $\gamma$) at one point but at *two* points or more: *bitangent*, or tritangent or... . But we can adapt the $A_k$ notation easily. A circle, with centre $u$ having 2-point contact ($A_1$ contact) with $\gamma$ at each of two separate places is said to have $A_1^2$ contact, or $A_1A_1$ contact, and $u$ can be called an $A_1^2$ point for $\gamma$. Likewise an $A_1A_2$ point $u$ is the centre of a circle having $A_1$ contact somewhere and $A_2$ contact somewhere else: the circle is the osculating circle at the latter point, which is not a vertex of $\gamma$. The centre $u$ of a "bi-osculating circle" is an $A_2^2$ point, or in fact an $A_{\geqslant 2}A_{\geqslant 2}$ point, being exactly $A_2^2$ when neither point of contact is a vertex of $\gamma$.

Suppose that a circle $C_0$, with centre $u_0$ and radius $r_0 > 0$, is bitangent to the simple curve $\gamma$ (simple: without self-intersections). We write $\gamma_1$ and $\gamma_2$ for small pieces of $\gamma$ close to the points of contact, and parametrize each $\gamma_i$ by arc-length $t_i$ in such a way that, for each $i$, $t_i = 0$ corresponds to the point of contact of $C_0$. Orient $\gamma_i$ as in Fig. 5, both anticlockwise around $C_0$. (It is possible that one or both of the $\gamma_i$ is oriented oppositely to $\gamma$; thus $\gamma_i(t_i) = \gamma(\pm t + k_i)$ for constants $k_i$ when $\gamma$ is parametrized by arc-length.) Let $T_i = \gamma_i'$ be the unit tangent to $\gamma_i$, and $N_i$ the unit normal, 90° anticlockwise from $T_i$.

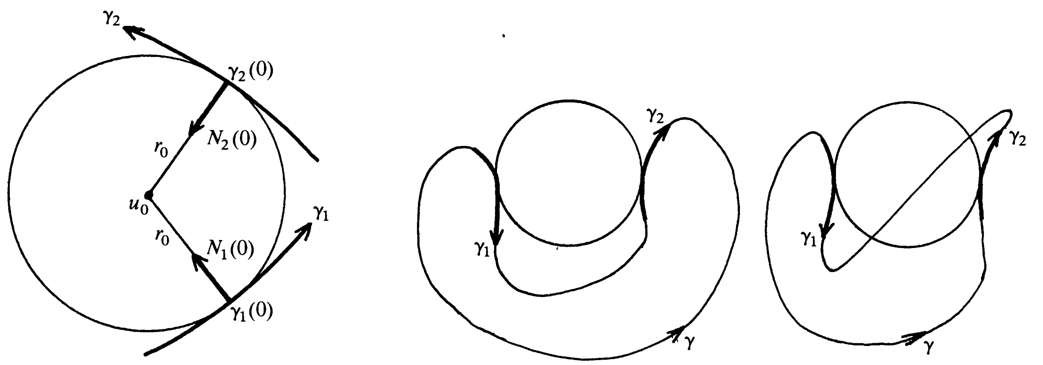



FIG. 5

Consider the function $g$: $\mathbb{R} \times \mathbb{R}, (0,0) \to \mathbb{R}, 0$ defined as follows. (The notation indicates that $g$ is defined on some neighbourhood of $(0,0)$ in the source $\mathbb{R} \times \mathbb{R}$, and that $g(0,0) = 0$.)

$$g(t_1, t_2) = (\gamma_1(t_1) - \gamma_2(t_2)) \cdot (T_1(t_1) - T_2(t_2)),$$

the dot being scalar product. Because of orientations, $T_1 - T_1$ cannot be zero.

Thus $g = 0$ if and only if $\gamma_1 - \gamma_2$ is perpendicular to $T_1 - T_2$,
if and only if $\gamma_1 - \gamma_2$ is parallel to $N_1 - N_2$,
if and only if $\gamma_1 - \gamma_2 + r(N_1 - N_2) = 0$, for some real number $r \neq 0$,
if and only if $\gamma_1 + rN_1 = \gamma_2 + rN_2$, for some $r \neq 0$.

When this happens, the common value $u$ of $\gamma_i + rN_i$ is the centre of a circle of radius $|r|$ touching $\gamma_1$ at $\gamma_1(t_1)$ and $\gamma_2$ at $\gamma_2(t_2)$. Thus $t_1 = t_2 = 0$ gives $u = u_0$ and $r = r_0$. Conversely, circles close to $C_0$ touching $\gamma_1$ at $\gamma_1(t_1)$ and $\gamma_2$ at $\gamma_2(t_2)$ will give

$$\gamma_1 + rN_1 = \gamma_2 + rN_2,$$

rather than, say

$$\gamma_1 + rN_1 = \gamma_2 - rN_2.$$

Thus $g = 0$ picks out just those pairs $(t_1, t_2)$ close to $(0,0)$ giving bitangent circles close to $C_0$.

In fact the computational technique used to plot the symmetry sets and midloci in this paper is precisely to look for the zeros of a function such as $g$, globally defined for a particular curve $\gamma$.

Thus we fix say $t_1$ and note changes of sign in $g(t_1, t_2)$ as $t_2$ varies.

The equation $g = 0$ itself defines a *curve* through the origin in the $(t_1, t_2)$ plane. If $\partial g/\partial t_2 \neq 0$ at (0, 0), then this curve is smooth, and $t_1$ can be used as a local coordinate on the curve; that is, $t_2$ is a smooth function of $t_1$, by the Implicit Function Theorem. The same holds with 1 and 2 interchanged. Using the standard formulas for plane curves (see e.g. **[3]**, Chap. 2 **[9]**), we find that, when $g = 0$, giving

$$\gamma_1 - \gamma_2 = -r(N_1 - N_2) \text{ at } (t_1, t_2),$$
$$\partial g/\partial t_i = (1 - T_1 \cdot T_2)(1 - r\kappa_i).$$

Now $1 - T_1 \cdot T_2 = 0$ if and only if $T_1 = T_2$ (unit vectors!), so the condition for $\partial g/\partial t_2 = 0$ is $r\kappa_2 = 1$, i.e., that the circle centre $u$ is the osculating circle at $\gamma_2(t_2)$, the radius $r$ being equal to the radius of curvature $1/\kappa_2$ at $t_2$.

Assuming that $\partial g/\partial t_2$ *is* nonzero at (0, 0), i.e., that $C_0$ is not the osculating circle at $\gamma_2(0)$, we can write $t_2 = t_2(t_1)$ on the curve $g^{-1}(0)$, the equation of this curve being $g = 0$. If $t_2' = dt_2/dt_1 \neq 0$ at $t_1 = 0$, then the curve $g^{-1}(0)$ is not tangent to the $t_1$-axis. Differentiating $g(t_1, t_2(t_1)) = 0$ with respect to $t_1$ we find that

$$t_2' = -(\partial g/\partial t_1)/(\partial g/\partial t_2).$$

Thus the condition $t_2' \neq 0$ becomes $1 \neq r_0\kappa_1(0)$, which says that $C_0$ is not the osculating circle at $\gamma_1(0)$ either. Examining $t_2''(0)$, which comes to $r_0\kappa_1'(0)/(1 - r_0\kappa_2(0))$ when $t_2'(0) = 0$, we find the following.

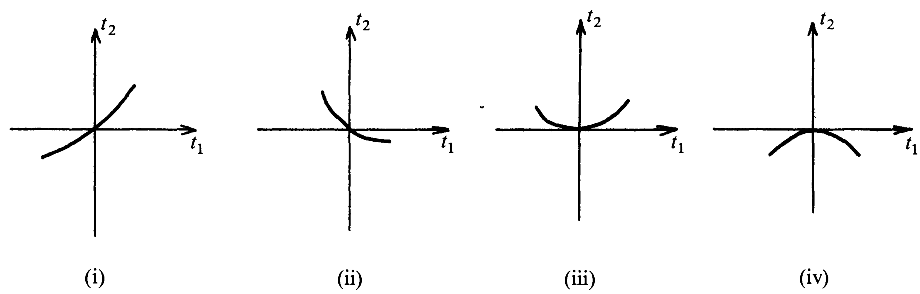



FIG. 6. The curve with equation $g = 0$.

(2.2) *Assume that the circle $C_0$ is not osculating at $\gamma_2(0)$, i.e., assume $1 \neq r_0\kappa_2(0)$. Then the local picture of $g^{-1}(0)$ is a smooth curve in the $(t_1, t_2)$ plane, as follows* (Fig. 6):

$A_1^2$ *contact*
- (i) *If $C_0$ is not osculating at $\gamma_1(0)$ and $1/r_0$ is between $\kappa_1(0)$ and $\kappa_2(0)$.*
- (ii) *If $C_0$ is not osculating at $\gamma_1(0)$ and $1/r_0$ is not between $\kappa_1(0)$ and $\kappa_2(0)$.*

$A_2A_1$ *contact*
- (iii) *If $C_0$ is osculating at $\gamma_1(0)$ and $\kappa_1'(0)\,(1 - r_0\kappa_2(0)) > 0$.*
- (iv) *If $C_0$ is osculating at $\gamma_1(0)$ and $\kappa_1'(0)\,(1 - r_0\kappa_2(0)) < 0$.*

This tells us how the points of contact move along $\gamma_1$ and $\gamma_2$, taking (say) $t_1$ to move steadily through 0 from negative to positive values. For instance, in case (iii) the point of contact $\gamma_2(t_2)$ moves from beyond $\gamma_2(0)$, back to $\gamma_2(0)$ and *then it turns round* (Fig. 7). Perhaps this leads us to

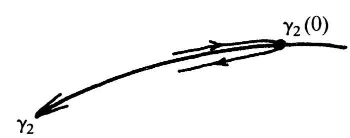



FIG. 7. The point of contact turns round.

expect some strange behaviour on the symmetry set or the midlocus. In fact, we shall see that there is (generally) a cusp on the symmetry set but that the midlocus scarcely notices the turnaround; see (6.2)(ii) and (3.1)(i) below.

What if the circle $C_0$ is bi-osculating, i.e., $1 = r_0\kappa_1(0) = r_0\kappa_2(0)$? Then the local picture of $g^{-1}(0)$ depends on the *second* partial derivatives of $g$ at $(0,0)$. In fact, calculation shows that the Taylor expansion of $g$ up to degree 2 terms is, when $1/r_0 = \kappa_1(0) = \kappa_2(0) = \kappa$, say,

$$g_2(t_1,t_2) = (\kappa_1'/\kappa)(N_1 \cdot N_2 - 1)t_1^2 + (\kappa_2'/\kappa)(N_1 \cdot N_2 - 1)t_2^2,$$

where the coefficients of $t_i^2$ are evaluated at $t_1 = t_2 = 0$. Now $N_1 \cdot N_2 \neq 1$, since $N_1(0) \neq N_2(0)$, so $g_2$ is a nondegenerate quadratic form provided $\kappa_1'(0), \kappa_2'(0)$ are both nonzero, i.e., provided neither $\gamma_1(0)$ nor $\gamma_2(0)$ is a *vertex*. Note that this says that the contact of $C_0$ with the curve $\gamma$ is exactly $A_2^2$. It is a standard fact of differential topology that a function, such as $g$, with Taylor expansion having no linear terms and nondegenerate quadratic terms (a "Morse function") can be reduced, by a change of variables in $\mathbb{R}^2$, to its quadratic part $g_2$. (See for example [7], p. 123.) That is to say there is a local diffeomorphism

$$\phi : \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$$

with

$$g(\phi(t_1,t_2)) = g_2(t_1,t_2);$$

the "higher terms" of $g$ have been *eliminated* by a mere change of variables. Since $\phi(g_2^{-1}(0)) = g^{-1}(0)$, the set $g^{-1}(0)$ is thrown onto $g_2^{-1}(0)$ by the diffeomorphism $\phi^{-1}$. This gives the following.


FIG. 8. Curve $g = 0$ in the bi-osculating case.

(2.3) *In the bi-osculating* ($A_2^2$) *case the local picture of* $g^{-1}(0)$ *is as follows* (Fig. 8):
(i) *an isolated point if* $\kappa_1'(0)\kappa_2'(0) > 0$,
(ii) *two smooth curves crossing at* $(0,0)$ *if* $\kappa_1'(0)\kappa_2'(0) < 0$.

Thus in case (i) there are *no* nearby pairs $(t_1, t_2)$ giving a bitangent circle, while in case (ii) there are two values of $t_2$ for each $t_1$ near 0 (and vice versa).

**3. The structure of the midpoint locus.** How can we use the information of §2 to describe the local structure of the symmetry set or the midlocus? For the midlocus it is easy. Let

$$m : \mathbb{R} \times \mathbb{R}, (0,0) \to \mathbb{R}^2$$

be defined by

$$m(t_1,t_2) = \frac{1}{2}(\gamma_1(t_1) + \gamma_2(t_2)).$$

(The notation indicates that $m$ is defined on some neighbourhood of $(0,0)$ in $\mathbb{R} \times \mathbb{R}$ but does not

specify $m(0,0)$.) Then $m(g^{-1}(0))$ is precisely the midlocus. If $m$ is actually a local diffeomorphism, then $g^{-1}(0)$ and the midlocus have exactly the same local structure up to a local diffeomorphism, that is up to a smooth change of coordinates in the plane. By the Inverse Function Theorem $m$ will be a local diffeomorphism if and only if the $2 \times 2$ Jacobian matrix of $m$ at $(0,0)$ is nonsingular, i.e., if and only if $T_1(0), T_2(0)$ are independent vectors. This can only *fail* if $T_1(0) = T_2(0)$, excluded above, or $T_1(0) = -T_2(0)$, Fig. 9.

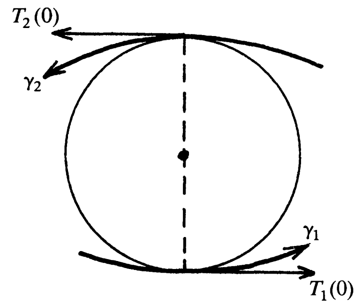



FIG. 9. $m$ is not a local diffeomorphism here.

Consider the $A_1^2$ and $A_1A_2$ cases, so that $g^{-1}(0)$ is smooth by (2.2). Then for the midlocus to be smooth it is not actually necessary that $m$ be a local diffeomorphism. We merely need *m restricted to* $g^{-1}(0)$ to give a smooth curve in $\mathbb{R}^2$, and this requires only that $m|g^{-1}(0)$ be an *immersion* at $(0,0)$, that is that its derivative at $(0,0)$ be nonzero. There are various ways of rephrasing this; for example parametrizing $g^{-1}(0)$ by $t_1$ we can require that

$$\delta(t_1) = \frac{1}{2}(\gamma_1(t_1) + \gamma_2(t_2(t_1)))$$

have $\delta'(0) \neq 0$. (Alternatively, Kernel $Dg(0,0)$ intersects Kernel $Dm(0,0)$ precisely in $(0,0)$, where $D$ stands for derivative.) This shows that $m|g^{-1}(0)$ is an immersion in the $A_1^2$ and $A_1A_2$ cases provided:

$$(*) \qquad T_1(0) \neq -T_2(0) \quad \text{or} \quad 2 \neq r_0(\kappa_1(0) + \kappa_2(0)).$$

When this is satisfied, the tangent direction to the curve $g^{-1}(0)$ in the $(t_1, t_2)$ plane is taken by $m$ (or more precisely by its derivative $Dm(0,0)$) to the tangent direction to the midlocus. We find:

(3.1) (i) *The midlocus is a smooth curve, for $(t_1, t_2)$ close to $(0,0)$, in the $A_1^2$ and $A_1A_2$ cases (that is, where $C_0$ is not bi-osculating), provided $(*)$ above holds. The tangent to the midlocus has direction $T_1(1 - r\kappa_2) - T_2(1 - r\kappa_1)$.*

(ii) *In the $A_2^2$ case (where $C_0$ is bi-osculating, but neither $\gamma_1(0)$ nor $\gamma_2(0)$ is a vertex) and assuming $T_1(0) \neq -T_2(0)$, the midlocus is locally an isolated point when $\kappa_1'(0)\kappa_2'(0) > 0$, and a pair of smooth crossing arcs when $\kappa_1'(0)\kappa_2'(0) < 0$. In the latter case with $\kappa_1'(0) = \alpha^2 > 0$, $\kappa_2'(0) = -\beta^2 < 0$, the tangent directions to the midlocus are $\beta T_1(0) \pm \alpha T_2(0)$.*

Before going on there is one more case for the midlocus worth mentioning. Suppose there is a straight *line* tangent to $\gamma_1$ at $\gamma_1(0)$ and $\gamma_2$ at $\gamma_2(0)$—see Fig. 10, where the curves could be on opposite sides of the line but are to be oriented left to right. Clearly there will be bitangent circles touching close to $\gamma_1(0)$ and $\gamma_2(0)$; the centres go off to infinity as the circles tend to the bitangent

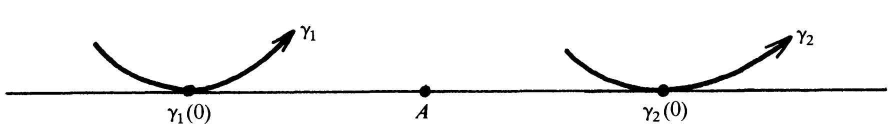



FIG. 10. Bitangent line.

line, but the midpoints of the chords of contact have for their limit the midpoint $A$ of the segment from $\gamma_1(0)$ to $\gamma_2(0)$. So we can ask for the local structure of the midlocus near $A$. Instead of $g$ we use the following function:

$$h: \mathbb{R} \times \mathbb{R}, (0,0) \to \mathbb{R}, 0,$$

where

$$h(t_1, t_2) = (\gamma_1(t_1) - \gamma_2(t_2)) \cdot (N_1(t_1) + N_2(t_2)).$$

Then, provided $N_1 + N_2 \neq 0$ (which is true at $t_1 = t_2 = 0$, and therefore nearby), $h(t_1, t_2) = 0$ if and only if there is a circle or line touching $\gamma_1$ at $\gamma_1(t_1)$ and $\gamma_2$ at $\gamma_2(t_2)$. Using this and similar arguments to the above we find the following.

(3.2) (i) *The curve* $h^{-1}(0)$ *in the* $(t_1, t_2)$ *plane is smooth at* $(0,0)$ *provided* $\kappa_1(0) \neq 0$ *or* $\kappa_2(0) \neq 0$, *i.e., provided the bitangent line is not bi-inflexional. The local picture of* $h^{-1}(0)$ *is shown in* Fig. 11, *where we assume* $\kappa_2(0) \neq 0$.

(ii) *The midlocus is smooth at A provided* $\gamma_1(0) \neq \gamma_2(0)$, *and its tangent direction is parallel to the bitangent line*. (See the right-hand half of Fig. 20.)

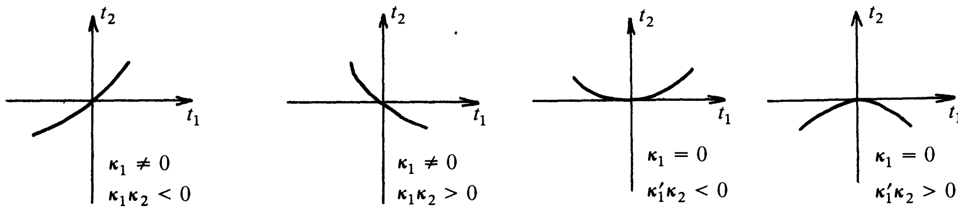



FIG. 11. The curve $h = 0$.

It is of course possible to find the local form of the midlocus as a curve parametrized by $t_1$. It comes to

$$\left(((\kappa_2 - \kappa_1)/2\kappa_2)t_1 + \text{higher terms}, (\kappa_1(\kappa_1 + \kappa_2)/2\kappa_2)t_1^2 + \text{higher terms}\right),$$

where $\kappa_1, \kappa_2$ are at 0. Thus when $\kappa_1 \neq 0$ and $\kappa_1 + \kappa_2 \neq 0$ (as well as $\kappa_1 \neq \kappa_2$) at $t_1 = t_2 = 0$ the local picture of the midlocus is as in Fig. 12, where we take $\kappa_2(0) > 0$. (When $\kappa_1(0) = 0$ the midlocus has an inflexion; can the reader discover which way it flexes?)

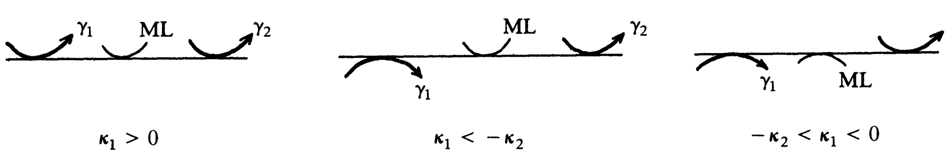



FIG. 12. Midlocus for a bitangent line.

(3.3) REMARK. There are various circumstances, detailed by 3.1 and 3.2, in which the midlocus is not smooth, namely: (i) the bi-osculating case $A_{\geqslant 2}A_{\geqslant 2}$, (ii) the $A_1^2$ and $A_1A_2$ cases when $T_1 = -T_2$ and $2 = r(\kappa_1 + \kappa_2)$, (iii) the bitangent line case when $\kappa_1 = \kappa_2$. All these are "non-generic" or *exceptional* in a precise sense (compare [4]). If a closed curve $\gamma$ has any two points where one of the above three things is true, then almost any sufficiently small perturbation of $\gamma$ will have *no* such points. So the midlocus is, for a "general" curve, always smooth (though it can have end-points and self-crossings). However, when we consider general *families* of curves, a non-smooth midlocus can very well occur for certain special curves of the family. We say something about non-smooth midloci in §5.

**4. The smooth part of the symmetry set.** We now turn to the symmetry set. This is not quite so easy, since it depends not only on the points $\gamma_1(t_1), \gamma_2(t_2)$ where a circle is tangent to two pieces

$\gamma_1, \gamma_2$ of $\gamma$ (as in Fig. 5), but also on the radius $r$. We start with a circle $C_0$ as in §2 above. Let

$$f\colon \mathbb{R} \times \mathbb{R} \times \mathbb{R}, (0,0,r_0) \to \mathbb{R}^2, 0$$

be defined by

$$f(t_1, t_2, r) = \gamma_1(t_1) - \gamma_2(t_2) + r(N_1(t_1) - N_2(t_2)).$$

Then $f(t_1, t_2, r) = 0$ precisely when a circle of radius $r$ is tangent to $\gamma_1$ at $\gamma_1(t_1)$ and $\gamma_2$ at $\gamma_2(t_2)$. We expect $f^{-1}(0)$ to be a *curve* in $\mathbb{R}^3$ (coordinates $(t_1, t_2, r)$), and the "centre" map

$$c\colon \mathbb{R}^3, (0,0,r_0) \to \mathbb{R}^2$$

be defined by

$$c(t_1, t_2, r) = \gamma_1(t_1) + rN_1(t_1)$$

takes $f^{-1}(0)$ to the symmetry set. Thus the symmetry set can be regarded as the "projection" $c(f^{-1}(0))$ by the map $c$ of the space curve given by $f = 0$ to the plane.

The Implicit Function Theorem can be applied to $f$ to find out when $f^{-1}(0)$ is a smooth curve near $(0,0,r_0)$: we require the 2 by 3 Jacobian matrix of $f$ to have rank 2 at $(0,0,r_0)$—that is, we require $f$ to be a *submersion* at $(0,0,r_0)$. (The Jacobian matrix of $f$ has column vectors $T_1(1-\kappa_1 r)$, $-T_2(1-\kappa_2 r)$, $N_1 - N_2$.) Then we can write down the condition for this curve to project under $c$ to a *smooth* curve in $\mathbb{R}^2$ (i.e., for $c|f^{-1}(0)$ to be an *immersion*): the tangent vector to $f^{-1}(0)$ at $(0,0,r_0)$ must not go to 0. (An equivalent condition is that Kernel $Df(0,0,r_0)$ intersects Kernel $Dc(0,0,r_0)$ in precisely $(0,0,0)$.)

(4.1) (i) *The space curve $f^{-1}(0)$ is a smooth curve parametrized by $t_1$ provided $1 \neq \kappa_2(0)r_0$, i.e., provided $C_0$ is not osculating at $\gamma_2(0)$. The same holds with 1 and 2 interchanged. Thus $f^{-1}(0)$ is smooth in the $A_1^2$ and $A_1A_2$ cases.*

(ii) *The symmetry set, namely $c(f^{-1}(0))$, is locally a smooth curve if $C_0$ is osculating at neither point $\gamma_1(0), \gamma_2(0)$ (the $A_1^2$ case). In that case either $t_1$ or $t_2$ can be used as a local parameter on the symmetry set. Its tangent vector is along the direction $T_1 - T_2$, the infinitesimal axis of symmetry* (Fig. 13).

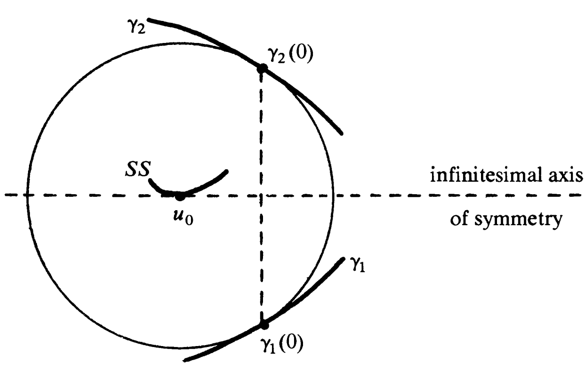



FIG. 13. Symmetry set touches the infinitesimal axis.

(4.2) REMARKS. (1) In the $A_1^2$ case it is not difficult to determine whether the symmetry set is concave upwards, as in Fig. 13, or downwards. For example, we can parametrize by $t_1$, writing say $\delta(t_1)$ for the symmetry set, and then the condition for upwards concavity is

$$\delta''(0) \cdot (N_1(0) - N_2(0)) > 0.$$

Calculation shows that this is equivalent to the following (all at $(0,0,r_0)$):

$\kappa_2 > \kappa_1$ when $1/r$ is between $\kappa_1$ and $\kappa_2$; $\kappa_2 < \kappa_1$ when $1/r$ is not between $\kappa_1$ and $\kappa_2$.

(Compare (2.2) (i) and (ii).) When $\kappa_1(0) = \kappa_2(0)$ the symmetry set has an *inflexion* at $u_0$, the centre of $C_0$. (Since by (4.1)(ii) the tangent to the symmetry set is the infinitesimal axis of symmetry, we can also consider the curve in the dual plane given by these axes—it is the dual of the symmetry set. It has a cusp when the symmetry set has an inflexion, that is when the curvatures agree at the points of contact, as above.) Note that, by (3.1)(i), the midlocus has its tangent along the infinitesimal axis of symmetry when $\kappa_1(0) = \kappa_2(0)$.

(2) At the cost of some complication we *can* eliminate $r$ and describe the symmetry set by $t_1$ and $t_2$ alone. For the equation $f = 0$ gives

$$r = (\gamma_1 - \gamma_2) \cdot (N_1 - N_2)/(N_1 \cdot N_2 - 1),$$

and this can be substituted in the formula for the map $c$.

**5. More on the midpoint locus.** We now turn to some of the cases (besides (3.1)(ii)) where the symmetry set or the midpoint locus is *not* smooth—even though, in the latter case, this happens only for nongeneric curves (but in generic families). First, here is an entertaining example.

(5.1) EXAMPLE. Consider the quartic curve $y = x^4 - ax^2 + bx$ (Fig. 14). Up to rigid motion in the plane every graph of a monic quartic polynomial has this form (can the reader prove this?). Provided $a > 0$ there is a bitangent line, namely $y = bx - \frac{1}{4}a^2$, touching the curve where $x = \pm(a/2)^{1/2}$. Is the midlocus smooth at the midpoint $A$ of the bitangent? We have two parameters $a$ and $b$ available—surely we can adjust them so that a smooth midlocus results? According to (3.2)(ii) we only have to check whether the curvatures of the quartic curve are different at the points where $x = \pm(a/2)^{1/2}$. Alas! the curvatures always coincide, both equalling $4a(1 + b^2)^{-3/2}$. So the midlocus is *never* smooth. (Actually, it is worse than that. See (5.4)(2) below, and the left-hand half of Fig. 18.)

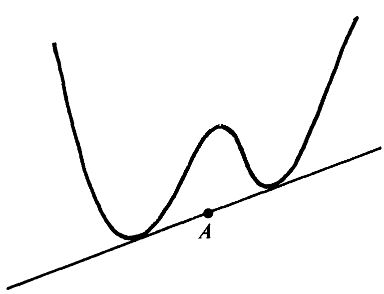



FIG. 14. Bitangent line of a quartic curve.

How can we get at the structure of a singular curve? In **[5]** we showed how to use the theory of unfoldings and discriminants to do just that for the symmetry set (see also **[3]** p. 137). However, the midlocus does not (so far as we know) arise in that way so that theory does not help with the midlocus. It has already been noted above (§3) that the map $m$ is a local diffeomorphism precisely when $T_1 \neq \pm T_2$. What if this condition fails at $(0,0)$? There is a whole hierarchy of maps $\mathbb{R}^2, 0 \to \mathbb{R}^2$ which are more degenerate than a local diffeomorphism; here we shall consider only one of these, the *fold map*; the *cusp map* or Whitney pleat is mentioned very briefly in (5.4) (2) below. The fold and cusp are of great significance in the study of maps from the plane to the plane: in fact singularity theory can be said to have begun with H. Whitney's proof in 1955 that, besides diffeomorphisms, these are the only *stable* maps—essentially unaltered by small perturbations. See for example **[6]** and **[10]**.

Let $\phi : \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ be defined by $\phi(t_1, t_2) = (t_1, t_2^2)$. This is a standard model ("normal form") for the fold map. To see what is going on it is helpful to factor the map through $\mathbb{R}^3$. Thus $\phi$ can be factored

$$\mathbb{R}^2 \xrightarrow{\phi_1} \mathbb{R}^3 \xrightarrow{\pi_1} \mathbb{R}^2$$

$$(t_1, t_2) \to (t_1, t_2^2, t_2)$$

$$(u_1, u_2, u_3) \to (u_1, u_2).$$

Here $\phi_1$ maps the plane into $\mathbb{R}^3$ as a surface and $\pi_1$ projects the surface "vertically" to the plane again (Fig. 15). The geometry of the fold map (and the cusp or Whitney pleat) is described in [**6**]. Given a map $\theta : \mathbb{R}^2, (0,0) \to \mathbb{R}^2$ it is possible to decide whether, by smooth changes of coordinates in the two copies of $\mathbb{R}^2$, the map $\theta$ can be converted into the normal form $\phi$ above. When this can be done we call $\theta$ a *fold map*. Because such a $\theta$ is "equivalent" to $\phi$ by merely changing coordinates, many properties of $\theta$ can be read off from the corresponding properties of $\phi$.

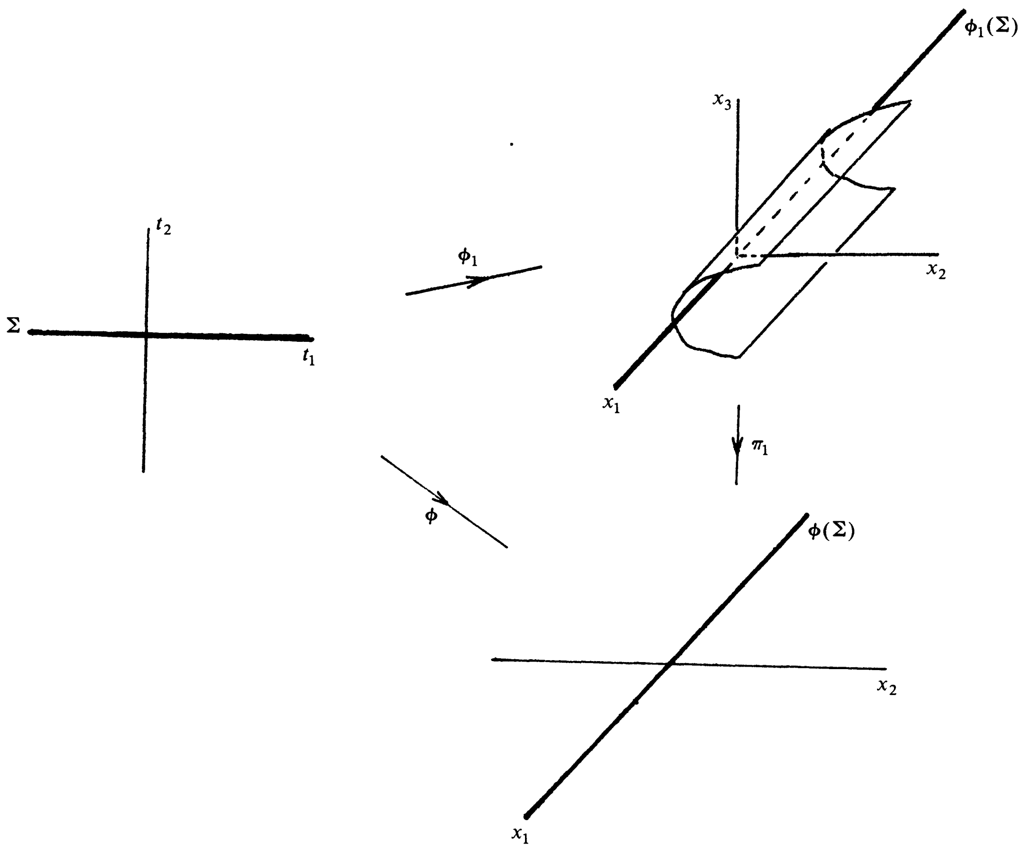



FIG. 15. Fold map.

There are two features of fold maps $\theta$ which are of special significance. The first is the *kernel line*, the line through $(0,0)$ in the direction of the tangent vector at $(0,0)$ sent to zero by $\theta$. For $\theta = \phi$ this is clearly the $t_2$-axis, in the direction $(0,1)$. For the map $m$ of §3 we need $T_1(0) = \pm T_2(0)$ for there to be such a line at all, and then it is the line of points $(t_1, t_2)$ satisfying $t_1 \pm t_2 = 0$, respectively. A curve through the origin whose tangent there happens to lie along the kernel line of a map $\theta$ will be sent by $\theta$ to a *singular* curve (Fig. 16,a) since its tangent vector at the origin will be collapsed to zero. Other curves will have (locally) smooth images under $\theta$ (Fig. 16,b). In terms of the pictures of Fig. 15 for $\phi$, the kernel line is in the direction which goes to a *vertical* direction by $\phi_1$. Curves with tangent vectors in the kernel direction go to curves which are

vertical at $(0,0,0)$, and, when the latter curves are projected by $\pi_1$, they become singular.

The other feature of interest to us is the *fold curve* $\Sigma$ in the source $\mathbb{R}^2$ of $\theta$. This consists of points $(t_1, t_2)$ at which the Jacobian matrix of $\theta$ is singular. For $\theta = \phi$, $\Sigma$ is the $t_1$-axis; for $\theta = m$ it is the set of $(t_1, t_2)$ for which $T_1(t_1) = \pm T_2(t_2)$. In the first case it is clear that $\Sigma$ is a smooth curve; for $\theta = m$ this needs proof. (To prove it consider the function $k(t_1, t_2) = T_1(t_1) \cdot N_2(t_2)$.

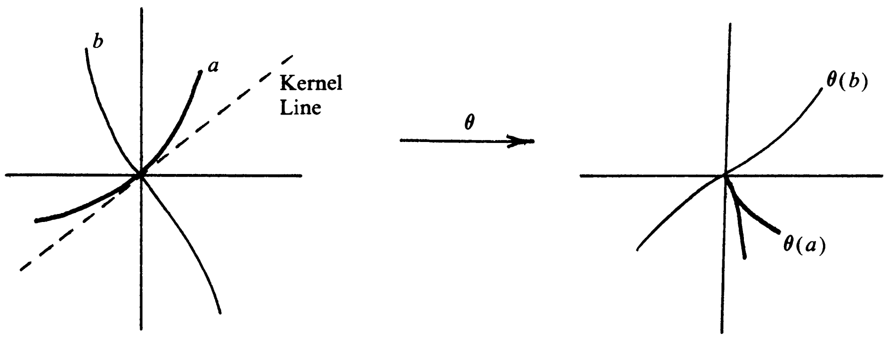



FIG. 16. Curves touching or not touching the kernel line.

It is easy to check that the derivative of $k$ is nonzero at $(0,0)$ provided $\kappa_1(0)$ and $\kappa_2(0)$ are not both zero; clearly $\Sigma = k^{-1}(0)$ so this shows $\Sigma$ is (locally) smooth provided we exclude the case $\kappa_1(0) = \kappa_2(0) = 0$. Note incidentally that $k(t_1, t_2) = T_1 \cdot T_2 - 1$ looks promising (or we might need $T_1 \cdot T_2 + 1$) but fails disastrously to prove $\Sigma = k^{-1}(0)$ is smooth!) The image $\theta(\Sigma)$ is of geometrical significance since it separates points $(x_1, x_2)$ in $\mathbb{R}^2$ which have different numbers of inverse images under $\theta$. For $\theta = \phi$, and hence for any fold map, the image $\theta(\Sigma)$ is a *smooth* curve which separates those $(x_1, x_2)$ with 0 inverse images from those with 2 (these have $x_2 < 0, x_2 > 0$, respectively).

For $\theta = m$ we need to know when $m$ *is* a fold map. We shall not go into the details of how this is checked (see [**8**], p. 146) but merely state the result.

(5.2) *The map $m$ is, at* $(0,0)$,
(i) *a local diffeomorphism if and only if* $T_1 \neq \pm T_2$;
(ii) *a fold map if and only if* $T_1 = \pm T_2$ *and* $\kappa_1 \neq \mp \kappa_2$, *respectively.*

*Here $T_i, \kappa_i$ are all at $t_1 = t_2 = 0$, of course. In* (ii) *the kernel line of $m$ is the line $t_1 = \mp t_2$ in $\mathbb{R}^2$.*

Let us look more closely at curves tangent to the kernel line at $(0,0)$ of the standard fold map $\phi$.

Consider in fact a curve $t_1 = a_2 t_2^2 + a_3 t_2^3 + \cdots$ in $\mathbb{R}^2$ tangent to the $t_2$-axis, which is the kernel line of $\phi$. The image under $\phi$ of this curve is the curve consisting of points

$$(u_1, u_2) = \left(a_2 t_2^2 + a_3 t_2^3 + \cdots, t_2^2\right).$$

A smooth change of coordinates $(u_1, u_2) \to (u_1 - a_2 u_2, u_2)$ turns this into the set of points $(a_3 t_2^3 + \cdots, t_2^2)$. Such a set of points in $\mathbb{R}^2$ is said to have an *ordinary cusp* (Fig. 17) at $(0,0)$ provided $a_3 \neq 0$. An ordinary cusp is the simplest kind of singularity a curve can have and, as the argument above shows, the standard fold map will take almost any smooth curve through $(0,0)$ to a curve with an ordinary cusp at the corresponding point. The latter result holds for any fold map, for such a map can be reduced to $\phi$ by changes of coordinates. (Formally, the condition for $\delta: \mathbb{R}, 0 \to \mathbb{R}^2$ to have an ordinary cusp at $\delta(0)$ is that $\delta(0) = 0$ and $\delta''(0), \delta'''(0)$ are linearly independent. This can be used to show that, if $\theta$ is equivalent to $\phi$ by smooth changes of coordinates, then an ordinary cusp occurs on $\theta \circ \delta$ if and only if it occurs on $\phi \circ \delta$.)

Applying this to the curve $g^{-1}(0)$ or $h^{-1}(0)$ and the map $m$ of §3, we find the following (compare (3.1), (3.2)).

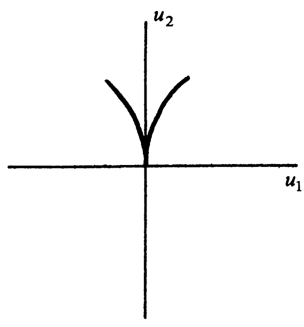



FIG. 17. Ordinary cusp.

(5.3) (i) *In the $A_1^2$ or $A_1A_2$ cases ($\kappa_1 r \neq 1$ or $\kappa_2 r \neq 1$), with $T_1 = -T_2$ and $2 = r_0(\kappa_1 + \kappa_2)$ at $t_1 = t_2 = 0$, the midlocus generally has an ordinary cusp and the map m is always a fold map.*

(ii) *Suppose there is a bitangent line which is not bi-inflexional ($\kappa_1 \neq 0$ or $\kappa_2 \neq 0$), but that $\kappa_1(0) = \kappa_2(0)$. Then the midlocus generally has an ordinary cusp and the map m is always a fold map.*

(5.4) REMARKS. (1) It is, of course, possible to find the precise conditions for an ordinary cusp, but they do not seem to be very illuminating. To convince the reader of this, here is the condition in (ii):

$$(\kappa_1' + \kappa_2')\bigl(4\kappa + 2\lambda(\kappa_1' - \kappa_2')\bigr) \neq \lambda\kappa(\kappa_1'' - \kappa_2'').$$

Here, as usual, primes denote derivatives with respect to $t_1$ or $t_2$ as appropriate, derivatives being evaluated at $t_1 = t_2 = 0$, and $\lambda$ is the length of the bitangent while $\kappa = \kappa_1(0) = \kappa_2(0)$. It is a curious fact that when we consider the case of the recalcitrant quartic curve (5.1), when the curvatures at the points of contact of the bitangent line are always equal, then, in fact, the non-equality condition above *always* fails: the midlocus *never* has an ordinary cusp! So from this point of view the family of quartic curves which are graphs of functions is extremely non-generic.

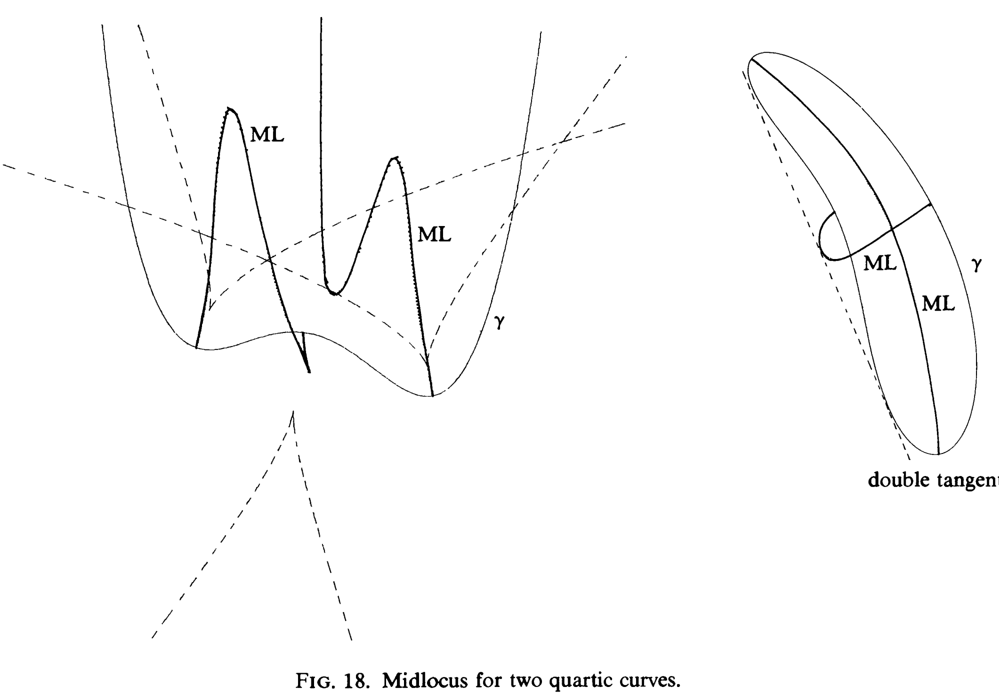



FIG. 18. Midlocus for two quartic curves.

See the left-hand half of Fig. 18. By contrast, the right-hand half shows a better behaved curve where the midlocus is smooth at the midpoint of the chord of contact of the double tangent. In fact this curve is also a quartic curve (it is an orthotomic of an ellipse) but not, of course, the graph of a function since it is a closed curve.

(2) For readers familiar with the cusp map (Whitney pleat) we mention that $m$ is a cusp map at $(0,0)$ if and only if $T_1 = \pm T_2$, $\kappa_1 = \mp\kappa_2 \neq 0$ and $\kappa_1' \neq \kappa_2'$ at 0. These conditions are not compatible with the hypotheses of (5.3).

Suppose that $T_1(0) = -T_2(0)$, but $\kappa_1(0) \neq \kappa_2(0)$, so that by (5.2)(ii) the map $m$ is a fold map. The fold curve $\Sigma$ is those $(t_1, t_2)$ for which $T_1(t_1) = -T_2(t_2)$, and $m(\Sigma)$ is the set of midpoints of the chords joining $\gamma_1(t_1)$ to $\gamma_2(t_2)$ for such $t_1, t_2$. Note that there are not necessarily any bitangent circles in sight at present; for that we need also $\gamma_1(t_1) - \gamma_2(t_2)$ to be perpendicular to $T_1(t_1)$. This locus of midpoints of chords joining points of parallel tangency is then a smooth curve and it separates points $(x_1, x_2)$ of the plane which are midpoints of *two* chords (joining points close to $t_1 = t_2 = 0$) from those which are midpoints of *none*. See Fig. 19.

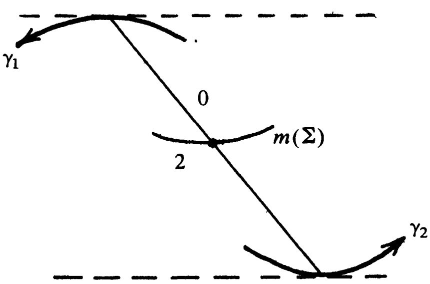



FIG. 19. Midpoints of chords of contact of parallel tangents.

**6. More on the symmetry set.** When we try to apply the ideas of fold maps to symmetry sets, we are faced with a difficulty. For the symmetry set is the image $c(f^{-1}(0))$, as in §4, where $c$ goes from $\mathbb{R}^3$ to $\mathbb{R}^2$. Where is there a map from $\mathbb{R}^2$ to $\mathbb{R}^2$? Let

$$\tilde{g}: \mathbb{R} \times \mathbb{R} \times \mathbb{R}, (0,0,r_0) \to \mathbb{R}$$

be defined by

$$\tilde{g}(t_1, t_2, r) = g(t_1, t_2) = (\gamma_1 - \gamma_2) \cdot (T_1 - T_2).$$

Then we have: if $f = 0$, then $\tilde{g} = 0$; hence

$$M = f^{-1}(0) \text{ is contained in } S = \tilde{g}^{-1}(0).$$

In general $S$ is a surface; when it is smooth and $M$ is a smooth curve in $S$ we can parametrize $S$ locally by two variables and so treat $S$ very much like a piece of $\mathbb{R}^2$. Then we are back to a curve in the plane.

In fact suppose $r_0\kappa_2(0) \neq 1$, while $f(0,0,r_0) = 0$ (thus the circle $C_0$ of radius $r_0$ is tangent to $\gamma_1$ and $\gamma_2$ at $t_1 = t_2 = 0$ and is not osculating at $t_1 = 0$). Applying the arguments leading to (2.2) to $\tilde{g}$ instead of to $g$, we can parametrize $S$ locally by $t_1$ and $r$. Then using $t_1$ and $r$ as local coordinates on $S$, and writing $\tilde{c}$ for the restriction of $c$ to $S$, we have:

$$\tilde{c}(t_1, r) = \gamma_1(t_1) + rN_1(t_1).$$

The curve $M$ can be thought of as a curve in the $(t_1, r)$ plane and $\tilde{c}(M)$ is the symmetry set (Fig. 20).

Notice that $\tilde{c}$ takes lines parallel to the $r$-axis to *normals* to $\gamma_1$, and the $t_1$-axis to the curve $\gamma_1$ itself. The fold set $\Sigma$ of $\tilde{c}$ is found by the condition that the 2 by 2 Jacobian matrix of $\tilde{c}$ is singular, i.e., that the columns

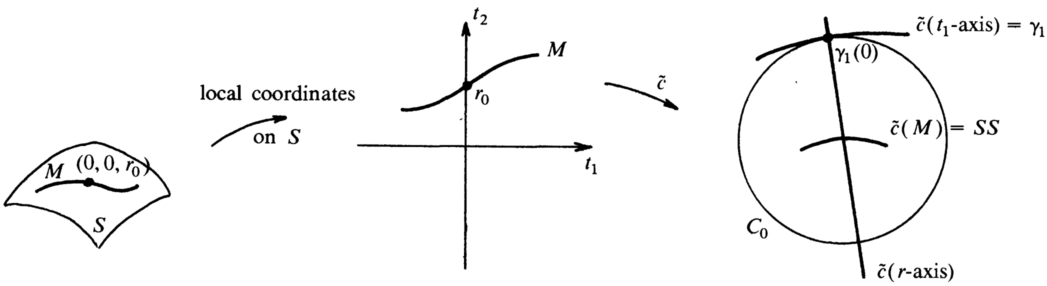



FIG. 20. Symmetry set via the map $\tilde{c}$.

$$T_1(1 - r\kappa_1) \quad \text{and} \quad N_1$$

are dependent. This requires $1 = r\kappa_1$, so

$$\Sigma = \{(t_1, r) : r\kappa_1(t_1) = 1\},$$

and $\tilde{c}(\Sigma)$ is precisely the *evolute* of $\gamma_1$. We find:

(6.1) *The map $\tilde{c}$ is, at $t_1 = 0$, $r = r_0$,*

(i) *a local diffeomorphism if and only if $\kappa_1 r \neq 1$ if and only if $C_0$ is not osculating at $t_1$ ($A_1^2$ case),*

(ii) *a fold map if and only if $\kappa_1 = 1$, $\kappa_1' \neq 0$ if and only if $C_0$ is osculating at $t_1$, not a vertex ($A_1A_2$).*

We now want to consider the symmetry set, which is $\tilde{c}(M)$. When does the tangent to $M$ lie along the kernel direction of $\tilde{c}$, assuming we are in case (ii) or (iii) above? The kernel direction is $(1,0)$ since at $(0, r_0)$ the Jacobian matrix of $\tilde{c}$ has columns $0, N_1(0)$. The tangent direction at $(0,0,r_0)$ to $M$ in $(t_1, t_2, r)$-space is $(1,0,0)$ and of course this corresponds exactly to the direction $(1,0)$ in the $(t_1, r)$ plane. Thus $M$ is always tangent to the kernel direction of $\tilde{c}$. This shows:

(6.2) *The symmetry set*

(i) *is smooth at an $A_1^2$ point (compare* (4.1) (ii));

(ii) *generally has an ordinary cusp at an $A_1A_2$ point.*

See Fig. 21, where we assume $\kappa_1(0)$ and $\kappa_1'(0)$ are $> 0$, and the evolute of $\gamma_1$ is drawn dashed. In fact, more delicate methods show that the symmetry set *always* has an ordinary cusp at an $A_1A_2$ point.

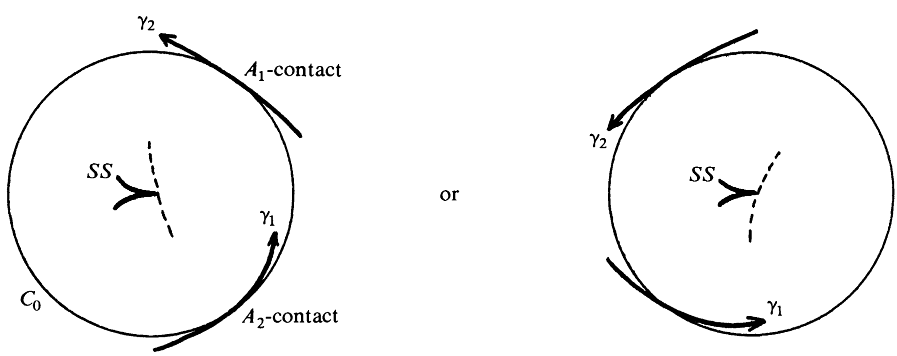



FIG. 21. Cusp on the symmetry set.

In the bi-osculating ($A_2^2$) case neither $M$ nor $S$ above is smooth. But in fact we can "recognize" $f$ to be a map rather like a fold map.

(6.3) *Assume that, at* $(0,0,r_0)$, *we have* $1 = r\kappa_1 = r\kappa_2$, $\kappa_1' \neq 0$, $\kappa_2' \neq 0$ ($A_2^2$ *case*). *Then by smooth changes of coordinates in* $\mathbb{R}^3$ *and* $\mathbb{R}^2$ *the map f can be written as*

$$(x_1, x_2, x_3) \to (x_1, x_2^2 \pm x_3^2),$$

*where the sign is* $+$ *or* $-$ *according as* $\kappa_1'(0)\kappa_2'(0)$ *is* $> 0$ *or* $< 0$.

Such an $f$ is a "generalized fold map", and from (6.3) we deduce that $M = f^{-1}(0)$ is, near $(0,0,r_0)$,

(i) an isolated point when $\kappa_1'(0)\kappa_2'(0) > 0$;
(ii) two smooth curves crossing when $\kappa_1'(0)\kappa_2'(0) < 0$.

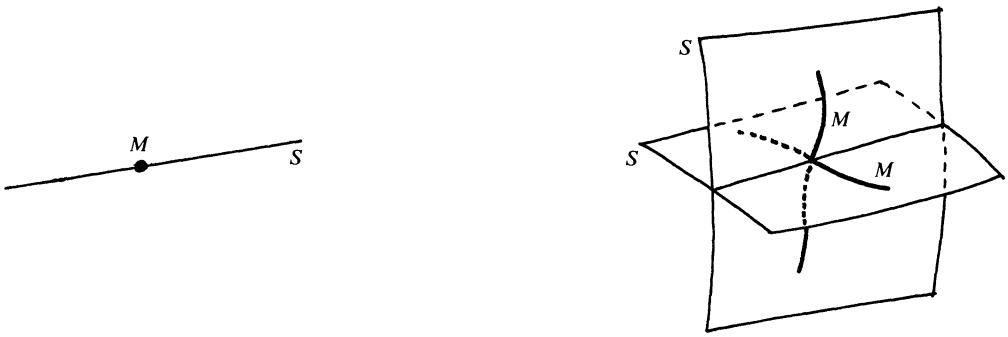



FIG. 22

In fact, $S = g^{-1}(0)$ is in these cases, respectively, a line (the $r$-axis) or a pair of smooth surfaces intersecting in the $r$-axis (compare (2.2)). See Fig. 22. In case (i) the symmetry set, which is $\tilde{c}(M)$, has an isolated point. By carefully analyzing case (ii) we find that the symmetry set consists of two cusps (generally ordinary cusps) pointing the *same* way, as in Fig. 23, where we take $\kappa_1, \kappa_2, \kappa_1'$ all $> 0$ and $\kappa_2' < 0$. The evolutes of $\gamma_1$ and $\gamma_2$ both pass through the centre of $C_0$ now; they are drawn dashed.

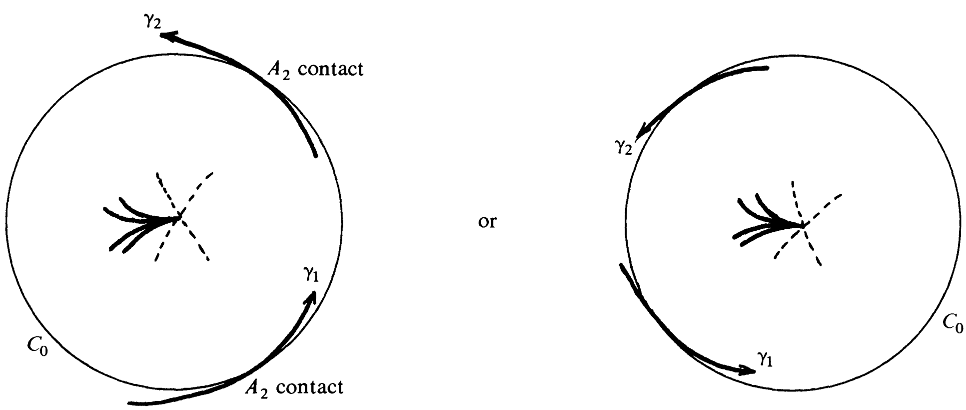



FIG. 23. Symmetry set: one bi-osculating case.

**7. Families of curves.** Finally we say a brief word about *families* of curves—that is, we suppose our starting curve $\gamma$ to be varying in a one-parameter family of smooth simple curves. Suppose we have a bi-osculating circle ($A_2^2$ case) as in (3.1)(ii) and (6.3). What effect will a small perturbation of the curves $\gamma_1$ and $\gamma_2$ have on the midlocus and the symmetry set? We merely draw

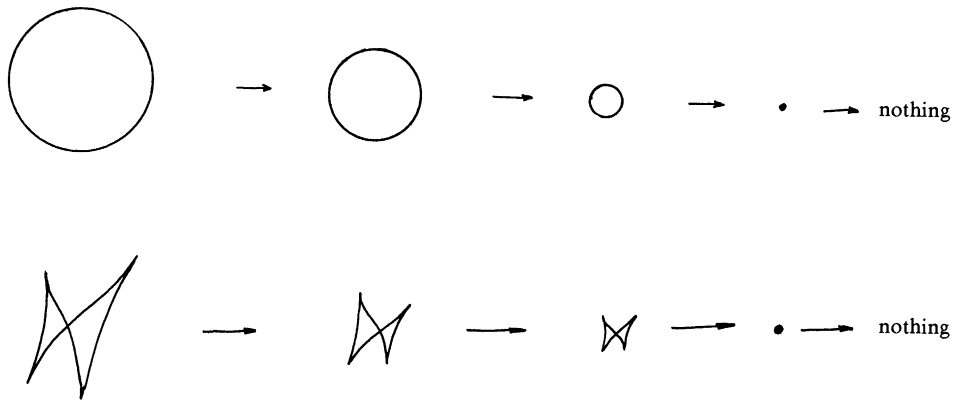



FIG. 24. Transitions on midloci (above) and symmetry sets (below) in one bi-osculating case.

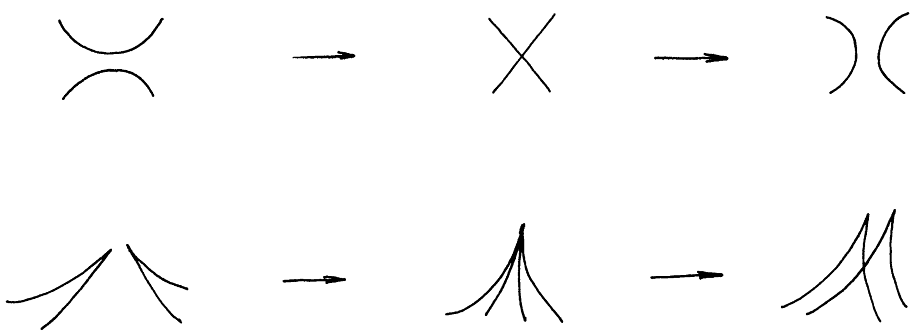


FIG. 25. More transitions: the other bi-osculating case.

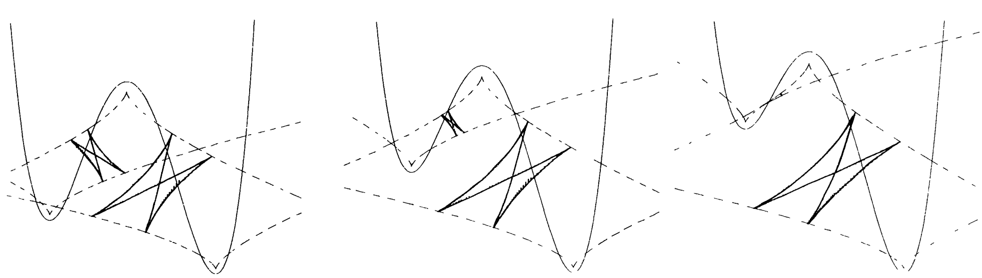


FIG. 26. A moth in the act of disappearing.

some pictures here, for a full discussion is beyond the scope of this article. In Fig. 24, top, a transition is shown on the midlocus whereby a closed curve shrinks and vanishes (the closed curve may not be simple—it could be more like a figure eight for example). In Fig. 24, bottom, we have what we call a "moth transition" on the symmetry sets: the moth shape shrivels to a point as in an intense candle flame and disappears! These assume $\kappa_1'\kappa_2' > 0$, as in (6.3)(i), which gives the isolated point in each picture. For $\kappa_1'\kappa_2' < 0$ the situation is illustrated in Fig. 25, the centre picture in each row coming from the $A_2^2$ case. Finally Fig. 26 shows an explicit example of a moth transition on a family of quartic curves. For clarity only part of the symmetry set is drawn in this diagram; as usual the evolute is drawn as a dashed curve.

**References**

1. H. Blum, Biological shape and visual science, I, J. Theoret. Biol., 38 (1973) 205–287.
2. M. Brady, Criteria for representations of shape, in Human and Machine Vision, Beck and Rosenfeld (eds.), Academic Press, 1983.
3. J. W. Bruce and P. J. Giblin, Curves and Singularities, Cambridge Univ. Press, 1984.
4. ______, Generic geometry, this MONTHLY, 90 (1983) 529–545.
5. J. W. Bruce, P. J. Giblin, and C. G. Gibson, Symmetry Sets, to appear in Proc. Roy. Soc. Edinburgh.
6. J. J. Callahan, Singularities of plane maps, this MONTHLY, 81 (1974) 211–240.
7. C. G. Gibson, Singular Points of Smooth Mappings, Research Notes in Mathematics 25, Pitman, 1979.
8. M. Golubitsky and V. Guillemin, Stable Mappings and their Singularities, Graduate Texts in Math. 14, Springer-Verlag, 1973.
9. E. Kreyszig, Differential Geometry, Univ. of Toronto Press, 1959.
10. T. Poston and I. Stewart, Catastrophe Theory and its Applications, Surveys and Reference Works in Math. 2, Pitman, 1978.
11. Y. Yomdin, On the local structure of a generic central set, Compositio Math., 43 (1981) 225–238.

# GEOMETRY WITHOUT POINTS

G. GERLA
*Dipartimento di Matematica e Applicazioni, Università degli Studi di Napoli, Italy*

R. VOLPE
*Via O. Morisani 13, Napoli, Italy*

A number of authors have addressed "the well-known conviction that geometry may be built without points" ([5]; see also the other references). In this paper we propose an approach to metric spaces, and thence to geometry, based on the primitives: solid, inclusion, and distance between solids.

DEFINITION. A *space of solids* is a set $S$ of elements called *solids*, together with a partial order relation $\leqslant$ on $S$, called *inclusion*, and a function $\delta$ from $S \times S$ to the nonnegative real numbers, called *distance*, such that, for all solids $u, v, w$:

(1) if $u \leqslant w$, then $\delta(u,v) \geqslant \delta(v,w)$,

(2) the *diameter* $\Delta(v) = \sup\{\delta(p,g) : p \leqslant v \text{ and } g \leqslant v\}$ is finite, and

$$\delta(u,v) + \delta(v,w) + \Delta(v) \geqslant \delta(u,w),$$

(3) $u \leqslant v$ if and only if $\delta(u',v) = 0$ for every $u'$ included in $u$,

(4) if $\varepsilon > 0$, then there are solids $u', v'$ such that

$$u' \leqslant u, \quad v' \leqslant v, \quad \Delta(u') \leqslant \varepsilon, \quad \Delta(v') \leqslant \varepsilon, \quad \text{and} \quad \delta(u',v') = \delta(u,v).$$

A *point* in a space of solids $(S, \leqslant, \delta)$ is a minimal element of $(S, \leqslant)$. An *$\varepsilon$-point*, where $\varepsilon > 0$, is a solid $u$ such that $\Delta(u) \leqslant \varepsilon$.

The following statements are easy consequences of the definition.

(a) For all $u$ and $v$, $\delta(u,v) = \delta(v,u)$.

G. Gerla received his degree in Mathematics from the University of Naples in 1970 and is now "Ricercatore" at the same University. His principal interests are in logic, epistemology, didactics, and the history of the mathematics in Naples (1600–1700).

R. Volpe received his degree in Mathematics from the University of Naples in 1981 with a thesis on modal logic. His principal interests have been in logic, epistemology, and the game of chess.

(b) The solid $u$ is a point if and only if $\Delta(u) = 0$.
(c) The set of all points in $S$ is a metric space with respect to $\delta$.
(d) If $u' \leqslant u$ and $v' \leqslant v$, then $\delta(u', v') \geqslant \delta(u, v)$.
(e) For all $u, u', v, v'$,

$$\delta(u,v) - \delta(u',v') \leqslant \delta(u,u') + \delta(v,v') + \Delta(u') + \Delta(v').$$

Hence if these solids are $\varepsilon$-points, then

$$|\delta(u,v) - \delta(u',v')| \leqslant \delta(u,u') + \delta(v,v') + 2\varepsilon.$$

Obviously a metric space $(M,d)$ is a space of solids with $S = M$, $\delta = d$, and $\leqslant$ simple equality. In this trivial example, every solid is a point. For a more interesting example obtained from $(M,d)$, let $S$ be the set of all nonempty compact subsets $u$ of $M$ such that $u^{\circ -} = u$, where $^{\circ}$ denotes interior and $^{-}$ denotes closure; define

$$\delta(u,v) = \min\{d(P,Q) : P \in u, Q \in v\},$$

and let $\leqslant$ be set inclusion. We show that $(S, \leqslant, \delta)$ is a space of solids, called the space *associated with* $(M,d)$.

Axiom (1) is clear. Since $v$ is compact, $\Delta(v) < \infty$. To complete the proof of (2), choose $U \in u$, $V_1 \in v$, $W \in w$, $V_2 \in v$ such that

$$\delta(u,v) = d(U,V_1) \text{ and } \delta(w,v) = d(W,V_2).$$

For $\varepsilon > 0$, let

$$D(V_1) = \{P \in v^{\circ} : d(P,V_1) < \varepsilon\} \text{ and } D(V_2) = \{P \in v^{\circ} : d(P,V_2) < \varepsilon\}.$$

Then $v_1 = D(V_1)^-$ and $v_2 = D(V_2)^-$ are solids contained in $v$. Choose $V_1' \in v_1$ and $V_2' \in v_2$ such that $\delta(v_1, v_2) = d(V_1', V_2')$. Then

$$d(V_1,V_2) \leqslant d\big(V_1,V_1'\big) + d\big(V_1',V_2'\big) + d\big(V_2',V_2\big) \leqslant \varepsilon + \delta(v_1,v_2) + \varepsilon$$

for all $\varepsilon > 0$; so $d(V_1,V_2) \leqslant \delta(v_1,v_2) \leqslant \Delta(v)$. Hence,

$$\delta(u,w) \leqslant d(U,W) \leqslant d(U,V_1) + d(V_1,V_2) + d(V_2,W) \leqslant \delta(u,v) + \delta(v,w) + \Delta(v).$$

For (3), it is obvious that $u \leqslant v$ implies $\delta(u',v) = 0$ for all $u' \leqslant u$. Conversely, suppose that $\delta(u',v) = 0$ for all $u' \leqslant u$ but that $u \nleqslant v$. Choose $P$ in $u^{\circ} \setminus v$ and $\varepsilon > 0$ such that $C = \{X \in M : d(X,P) \leqslant \varepsilon\}$ is disjoint from $v$. Let $B = \{X \in u^{\circ} : d(X,P) < \varepsilon\}$. Then $u' = B^-$ is a solid and $u' \cap v = \varnothing$. Then $\delta(u',v) > 0$, a contradiction; hence, $u \leqslant v$.

For (4), given $u$, $v$, and $\varepsilon$, choose $U$ and $V$ such that $\delta(u,v) = d(U,V)$, let $D(U) = \{X \in u^{\circ} : d(X,U) < \varepsilon\}$ and $D(V) = \{X \in v^{\circ} : d(X,V) < \varepsilon\}$. Then $u' = D(U)^-$ and $v' = D(V)^-$ are $\varepsilon$-points and $\delta(u',v') = \delta(u,v)$.

Recall that a metric space $(M,d)$ is *perfect* if it has no isolated points.

It is easy to see that

(f) a metric space is perfect if and only if its associated space of solids has no points.

The association of a space of solids with a metric space is characteristic of all spaces of solids, in the following sense.

THEOREM. *Every space of solids* $(S, \leqslant, \delta)$ *is isomorphic to a space of solids* $(S', \subseteq, \delta')$, *where* $S'$ *is a family of nonempty subsets of a metric space* $(M,d)$ *and* $\delta'[u,v] = \inf\{d(P,Q) : P \in u$ *and* $Q \in v\}$.

*Proof.* Let $\overline{M}$ be the set of all sequences $\overline{P} = (p_n)$ of solids such that $p_{n+1} \leqslant p_n$ for all $n$ and $\lim \Delta(p_n) = 0$. For $\overline{P}$ and $\overline{Q} = (q_n)$ in $\overline{M}$, define $\bar{d}(\overline{P},\overline{Q}) = \lim \delta(p_n, q_n)$. Since $(\delta(p_n,q_n))$ is increasing and bounded above by $\delta(p_1,q_1) + \Delta(p_1) + \Delta(q_1)$, $\bar{d}$ is defined on $\overline{M} \times \overline{M}$. It is easy to show that $(\overline{M}, \bar{d})$ is a pseudometric space. Let $M = \overline{M}/\mathscr{R}$, where $\overline{P}\mathscr{R}\overline{Q}$ means $\bar{d}(\overline{P},\overline{Q}) = 0$,

and for $P, Q \in M$ let $d(p, Q) = \bar{d}(\bar{P}, \bar{Q})$, where $\bar{P}$ is in the equivalence class $P$ and $\bar{Q}$ is in the equivalence class $Q$. Then $(M, d)$ is a metric space, called the space *associated with* $(S, \leqslant, \delta)$. Define the relation *belongs to*, $\beta \subseteq M \times S$, as follows: $P\beta u$ means there is a sequence $(p_n)$ in $P$ with $p_1 = u$. Define a function $f$ from $S$ into the set of all subsets of $M$ by

$$f(u) = \{P \in M: P\beta u\}.$$

Let $S'$ be the image of $f$. We must prove, for all $u, v$ in $S$: (i) $f(u) \neq \varnothing$, (ii) $u \leqslant v \Rightarrow f(u) \subseteq f(v)$, (iii) $\delta(u, v) = \delta'[f(u), f(v)]$, (iv) $f$ is injective.

(i) By repeated use of (4) we may build a sequence $(p_n)$ such that $p_1 = u$, $p_{n+1} \leqslant p_n$, and $\Delta(p_n) = 1/(n+1)$. Then the equivalence class of this sequence is in $f(u)$.

(ii) If $u \leqslant v$, then any sequence $u = p_1 \geqslant p_2 \geqslant \cdots$ may be extended to $v \geqslant u \geqslant p_2 \geqslant \cdots$.

(iii) Suppose $P\beta u$ and $Q\beta v$. Choose sequences $\bar{P} = (p_n)$ and $\bar{Q} = (q_n)$ with $p_1 = u$, $q_1 = v$. Then

$$\delta(u, v) = \delta(p_1, q_1) \leqslant \bar{d}(\bar{P}, \bar{Q}) = d(P, Q);$$

so

$$\delta(u, v) \leqslant \inf\{d(P, Q) : P\beta u \text{ and } Q\beta v\} = \delta'[f(u), f(v)].$$

Now by (4) we may build $(p_n'), (q_n')$ such that

$$p_1' = u,\ q_1' = v,\ p_{n+1}' \leqslant p_n',\ q_{n+1}' \leqslant q_n',\ \Delta(p_n') \leqslant 1/n,\ \Delta(q_n') \leqslant 1/n,$$

and

$$\delta(p_{n+1}', q_{n+1}') = \delta(p_n', q_n') = \delta(u, v).$$

Then these sequences represent equivalence classes $P', Q'$ such that $\delta(u, v) = d(P', Q')$. Thus

$$\delta(u, v) \geqslant \inf\{d(P, Q) : P\beta u \text{ and } Q\beta v\}.$$

Hence $\delta(u, v) = \delta'[f(u), f(v)]$.

(iv) If $u \neq v$, then we may assume $u \nleqslant v$. By (3) there exists $u' \leqslant u$ such that $\delta(u', v) > 0$. Then by (iii),

$$\inf\{d(P, Q) : P\beta u' \text{ and } Q\beta v\} > 0;$$

hence $f(u') \nleqslant f(v)$. But by (ii), $f(u') \leqslant f(u)$; hence $f(u) \neq f(v)$.

FURTHER REMARKS. Since euclidean, elliptic, and hyperbolic geometries may be defined solely in terms of axioms about metric spaces [**1**], the definition above leads to axiomatizations of geometries without the primitives "point". It remains to reformulate the resulting axioms in a more natural, less convoluted way. As one example, define three $\varepsilon$-points $u, v, w$ in a space of solids $(S, \leqslant, \delta)$ to be *$\varepsilon$-collinear* if *one* of the following holds:

$$\delta(u, v) + \delta(v, w) - \delta(u, w) < 3\varepsilon,$$

$$\delta(u, w) + \delta(w, v) - \delta(u, v) < 3\varepsilon,$$

$$\delta(v, u) + \delta(u, w) - \delta(v, w) < 3\varepsilon.$$

Then the axiom

(A) *there exist three noncollinear points*

holds in the associated metric space $(M, d)$ if and only if the axiom

(A′) *for some positive $\varepsilon$ there exist three $\varepsilon$-points that are not $\varepsilon$-collinear*,

holds in $(S, \leqslant, \delta)$.

*Proof*. Suppose that $P$, $Q$ and $T$ are noncollinear elements of $M$; let $(p_n)$, $(q_n)$ and $(t_n)$ be in $P$, $Q$ and $T$. Then there exists $\varepsilon > 0$ such that

$$d(P,T)+d(T,Q)-d(P,Q)>3\varepsilon,\ d(P,Q)+d(Q,T)-d(P,T)>3\varepsilon$$

and

$$d(Q,P)+d(P,T)-d(Q,T)>3\varepsilon.$$

It follows that there exists $m \in N$ such that

$$\Delta(p_n)<\varepsilon,\ \Delta(q_n)<\varepsilon,\ \Delta(t_n)<\varepsilon,\ \delta(p_n,q_n)+\delta(q_n,t_n)-\delta(p_n,t_n)>3\varepsilon,$$
$$\delta(p_n,t_n)+\delta(t_n,q_n)-\delta(p_n,q_n)>3\varepsilon \text{ and } \delta(q_n,p_n)+\delta(p_n,t_n)-\delta(q_n,t_n)>3\varepsilon$$

for every $n \geqslant m$. This proves that $p_m$, $q_m$ and $t_m$ are not $\varepsilon$-collinear $\varepsilon$-points.

Conversely, suppose that there exist three $\varepsilon$-points $u$, $v$ and $w$ that are not $\varepsilon$-collinear. Thus, if $P\beta u$, $Q\beta v$ and $T\beta w$, we have that

$$d(P,T)+d(T,Q)-d(P,Q)>0.$$

Indeed, observe that

$$\delta(u,w)+\delta(w,v)-\delta(u,v)-2\varepsilon \geqslant \varepsilon.$$

Now, by (iii), there exist $U\beta u$, $U'\beta u$, $V\beta v$, $V'\beta v$, $W\beta w$, $W'\beta w$ such that

$$\delta(u,w)=d(U,W),\ \delta(w,v)=d(W',V),\ \delta(u,v)=d(U',V').$$

Then

$$\begin{aligned} d(P,T)+d(T,Q)-d(P,Q) &\geqslant \delta(u,w)+\delta(w,v)-d(P,Q)\\ &\geqslant \delta(u,w)+\delta(w,v)-[d(P,U')+d(U',V')+d(V',Q)]\\ &\geqslant \delta(u,w)+\delta(w,v)-[\varepsilon+\delta(u,v)+\varepsilon]=\delta(u,w)+\delta(w,v)-\delta(u,v)-2\varepsilon \geqslant \varepsilon>0. \end{aligned}$$

In the same manner one proves that

$$d(P,Q)+d(Q,T)-d(P,T)>0$$

and that

$$d(Q,P)+d(P,T)-d(Q,T)>0.$$

In conclusion, $P$, $Q$ and $T$ are noncollinear.

Axiom A can also be reformulated with no mention of points at all. Define three solids $u, v, w$ to be *collinear* if

$$\delta(u,v)+\delta(v,w)-\delta(u,w)<3m$$

or

$$\delta(u,w)+\delta(w,v)-\delta(u,v)<3m$$

or

$$\delta(v,u)+\delta(u,w)-\delta(v,w)<3m,$$

where $m=\max[\Delta(u),\Delta(v),\Delta(w)]$. Then A holds in $(M,d)$ if and only if

(A″) *there exist three noncollinear solids*

holds in $(S, \leqslant, \delta)$.

**References**

1. L. M. Blumenthal, A Modern View of Geometry, Freeman, San Francisco and London, 1961.

2. B. L. Clarke, A calculus of individuals based on connection, Notre Dame J. Formal Logic, vol. 22, n. 3 (1982) 204–218.

3. T. De Laguna, Point, line and surface, as sets of solids, J. Philosophy, 19 (1922) 449–461.

4. G. Gerla, La geometria senza punti, Archimede, 3 (1982) 139–146.

5. A. Grzegorczyk, Axiomatizability of geometry without points, Proceedings of the colloquium at Utrecht, January, 1960, Dordrecht, 1961, 104–111.

6. E. V. Huntington, A set of postulates for abstract geometry exposed in terms of the simple relation of inclusion, Math. Ann. lxxiii (1916) 522–529.

7. S. Jaskowski, Une modification des définitions fondamentales, de la géometrie des corps de M. A. Tarski, Ann. Soc. Math. Polon., XXI (1948) 298–301.

8. J. Nicod, La géometrie dans le monde sensible, Paris, 1924.

9. T. F. Sullivan, Affine geometry having a solid as primitive, Notre Dame J. Formal Logic, XII (1971) 1–61.

10. A. Tarski, Foundation of geometry of solids, in Logic Semantics Metamathematics, transl. J. H. Woodger, Oxford (1956) 24–30.

# CENTROSYMMETRIC (CROSS-SYMMETRIC) MATRICES, THEIR BASIC PROPERTIES, EIGENVALUES, AND EIGENVECTORS

JAMES R. WEAVER

*Department of Mathematics / Statistics, The University of West Florida, Pensacola, FL 32514*

**1. Introduction.** For a number of decades symmetric matrices over the real field have been studied intently by every beginning linear algebra student as a class of matrices which are defined by the property of being symmetric about their main diagonal or by the fact that you can interchange the rows and columns of a given symmetric matrix and it remains unchanged. In [**1**] (p. 124) A. C. Aitken defines a centrosymmetric matrix $P$ and this definition coincides with the definition given by Graybill of a cross-symmetric matrix [**6**] (p. 361).

DEFINITION 1. An $n \times n$ matrix $P$ over the real field is centrosymmetric if

$$P_{i,j} = P_{n-i+1,n-j+1}, \quad \text{for } 1 \leqslant i, j \leqslant n.$$

A closer look at Definition 1 reveals that a centrosymmetric matrix is nothing more than a square matrix which is symmetric about its center.

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & b & c \\ d & e & d \\ c & b & a \end{pmatrix}$$

are examples of centrosymmetric matrices. This type of matrix is very important in the study of certain types of Markov Processes because they turn out to be transition matrices for the process. As we look at centrosymmetric matrices we will find that they have many interesting properties, comparable in some ways with symmetric matrices, but yet in some ways they are very different from symmetric matrices. For instance; it will be shown that they form an algebra.

**2. Historical Notes.** M. Iosifescu points out in [**7**] that "The concept of Markov dependence appears in an explicit form in 1906 in a paper [**12**] of the Russian mathematician A. A. Markov (1856–1922). In a series of papers starting with [**12**] he studied various properties of sequences of dependent random variables which in his honor are nowadays called finite Markov chains.

Almost at the same time, studying the problem of shuffling the card deck, the French

Jim Weaver was born in Columbus, Ohio, in 1941 and did his undergraduate work in mathematics at Kent State University in Ohio. He wrote his dissertation in group theory under Joseph Adney at Michigan State University. Since then he has been at The University of West Florida, except for a sabbatical leave which was spent at The University of Massachusetts in Amherst, Massachusetts. His primary interest at this time is in linear algebra and he has been intimately involved with mathematics competition for the community colleges of the southeastern portion of the United States since 1973.

mathematician Henri Poincaré came across sequences of random variables which are in fact Markov chains. Nevertheless, Poincaré did not undertake a systematic study of those sequences so that the recognition accorded to Markov is entirely justified."

S. Wright [15] introduced the idea of random genetic drift in 1931 and the mathematical model of this physical phenomenon is a Markov process and was developed by G. Malécot [11] in 1944. M. Iosifescu [7] examines a special case of this phenomenon, which consists of a population consisting of individuals that are able to produce a single type of gamete. The transition matrix $Q$ for this Markov chain is an $(n+1)\times(n+1)$ centrosymmetric matrix where

$$Q_{i+1,j+1}=\binom{n}{j}\left(\frac{n-i}{n}\right)^{n-j}\left(\frac{i}{n}\right)^{j},\qquad 0\leqslant i,j\leqslant n.$$

It should be noted that the entries of this transition matrix are binomial probabilities. R. G. Khazanie [8], R. G. Khazanie and H. E. McKean [9], and F. Barnett, J. S. Frame, and J. Weaver [3] give an explicit formula for the eigenvalues of this matrix and in [3] an explicit formula for the eigenvectors is found.

M. Kimura [10] studied a Markov process in his article "Some Problems of Stochastic Processes in Genetics" whose transition matrix $R$ is an $(n+1)\times(n+1)$ centrosymmetric matrix. The matrix $R$ is given when $\alpha=2$ by

$$R_{i+1,j+1}=\binom{(n-i)\alpha}{n-j}\binom{i\alpha}{j}\Big/\binom{n\alpha}{n},\qquad 0\leqslant i,j\leqslant n.$$

It should be noted that the entries of $R$ are hypergeometric probabilities.

I. Schensted [13] studied the same centrosymmetric matrix $R$ as M. Kimura [10] in her paper entitled "Model of Subnuclear Segregation in the Macronucleus of Ciliates". A formula for the eigenvalues of this centrosymmetric matrix is given when $\alpha=2$ by both Kimura and Schensted but neither gives a formula for the eigenvectors. In [14] J. Weaver finds a formula for the eigenvalues and eigenvectors of $R$ for any positive integer $\alpha>1$. For $\alpha=1$, $\mathbf{R}=\mathbf{I}$.

I. J. Good points out in [5] that a symmetric Toeplitz matrix is an example of a matrix which is both symmetric and centrosymmetric. He also states in [5] that "Toeplitz matrices arise as discrete approximations to kernels $k(x,t)$ of integral equations when these kernels are functions of $|x-t|$. Similarly if a kernel is an even function of its vector argument $(x,t)$, that is $k(x,t)=k(-x,-t)$, then it can be discretely approximated by a centrosymmetric matrix."

With this historical background as a backdrop, we pursue a more detailed study of centrosymmetric matrices.

**3. Eigenvalues and Eigenvectors of Small Centrosymmetric Matrices with Their Row Sums Equal.** Note that the $2\times 2$ centrosymmetric matrix $\begin{pmatrix} a & b\\ b & a\end{pmatrix}$ has $a+b$ and $a-b$ as its eigenvalues with eigenvectors $(1,1)'$ and $(1,-1)'$, respectively. The centrosymmetric matrix $\begin{pmatrix} a & b & c\\ d & e & d\\ c & b & a\end{pmatrix}$ whose row sums are equal has eigenvalues $a+b+c$, $a-c$, and $e-b$ with eigenvectors $(1,1,1)'$, $(1,0,-1)'$, and $(b,-2d,b)'$, respectively, if $b$ or $d$ is not equal to zero. It is regrettable, but these nice formulas cannot be extended to the $4\times 4$ case.

Once some basic information about centrosymmetric matrices is obtained, some theorems are given which are very useful when it comes to finding the eigenvalues and eigenvectors of any centrosymmetric matrix.

**4. Basic Properties of Centrosymmetric Matrices.** As opposed to the definition of centrosymmetric matrices given by A. C. Aitken [1] and F. Graybill [6], a slightly different definition will now be given.

DEFINITION 2. If $P$ is an $m\times n$ matrix, then it is centrosymmetric if

$$P_{i,j}=P_{m-i+1,n-j+1}\quad\text{for } 1\leqslant i\leqslant m\quad\text{and}\quad 1\leqslant j\leqslant n.$$

It is not unusual to talk about a symmetric matrix as a matrix which is equal to its transpose. After the "reflection" of a matrix is defined we will be able to say a matrix is centrosymmetric if it is equal to its reflection.

DEFINITION 3. If $P$ is a $m \times n$ matrix, then the reflection of $P$ is $P^R$ and

$$[P^R]_{i,j} = P_{m-i+1,n-j+1} \quad \text{for } 1 \leqslant i \leqslant m \quad \text{and} \quad 1 \leqslant j \leqslant n.$$

If $P = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$, then $P^R = \begin{pmatrix} f & e & d \\ c & b & a \end{pmatrix}$ and an immediate consequence of Definition 3 is

PROPOSITION 4. *An $m \times n$ matrix $P$ is centrosymmetric if and only if $P^R = P$.*

One observation that simplifies work with centrosymmetric matrices is given once the matrix $J$ is defined.

DEFINITION 5. Let $J$ be the $n \times n$ symmetric elementary matrix defined by $J_{i,j} = \delta_{i,n-j+1}$, for $1 \leqslant i, j \leqslant n$, where $\delta_{i,j}$ is the Kronecker delta.

Note that $J$ has ones along the secondary diagonal and zeros elsewhere. The order of $J$ will differ in what is to follow, but its order will usually be clear from the context in which it is used. $J$ is an orthogonal matrix which is symmetric and hence $J^2 = I$, where $I$ is the identity matrix. Premultiplication of a matrix by $J$ will reverse the order of the rows and postmultiplication of a matrix by $J$ will reverse the order of the columns. Using the matrix $J$ one obtains

PROPOSITION 6. *An $m \times n$ matrix $P$ is centrosymmetric if and only if $JP = PJ$.*

*Proof.* If $JP = PJ$, then $P = JPJ$ and

$$P_{i,j} = [JPJ]_{i,j} = [PJ]_{m-i+1,j} = P_{m-i+1,n-j+1} \quad \text{for } 1 \leqslant i \leqslant m \quad \text{and} \quad 1 \leqslant j \leqslant n.$$

Therefore $P$ is centrosymmetric.

Conversely, if $P$ is centrosymmetric, then it is easy to show that $JP = PJ$.

Using Proposition 6, one is able to obtain

THEOREM 7. *The set of $n \times n$ real centrosymmetric matrices form an algebra.*

*Proof.* If $P$ and $Q$ are $n \times n$ real centrosymmetric matrices, then

$$J(P+Q)J = P + Q, \quad JPQJ = PQ,$$

and $J(cP)J = cP$ for a scalar $c$. It then follows that the set of $n \times n$ real centrosymmetric matrices form an algebra.

Two easy consequences of Proposition 6 are that if $P$ is a nonsingular $n \times n$ centrosymmetric matrix, then $P^{-1}$ is a centrosymmetric matrix, and if $P$ is an $n \times n$ centrosymmetric matrix, then $P^t$ is also a centrosymmetric matrix.

In **[5]** I. J. Good shows how to find the inverse of an $n \times n$ nonsingular centrosymmetric matrix and in **[1]** A. C. Aitken shows how to find the determinant of an $n \times n$ centrosymmetric matrix. In **[4]** A. Cantoni and P. Butler present information about $n \times n$ matrices which are symmetric and centrosymmetric. Many of the results established by A. C. Aitken **[1]**, A. Cantoni and P. Butler **[4]**, and I. J. Good **[5]** make extensive use of the following orthogonal transformation $Q$.

DEFINITION 8. If $n = 2s$ is a positive even integer, then

$$Q = \sqrt{\tfrac{1}{2}} \begin{pmatrix} I & -J \\ J & I \end{pmatrix},$$

and if $n = 2s + 1$ is a positive odd integer, then

$$Q = \sqrt{\tfrac{1}{2}} \begin{pmatrix} I & 0 & -J \\ 0 & \sqrt{2} & 0 \\ J & 0 & I \end{pmatrix},$$

where $I$ and $J$ are $s \times s$ matrices.

In either case it is easy to see that $Q$ is an $n \times n$ orthogonal matrix which diagonalizes $J$. Using the appropriate $Q$ on an $n \times n$ centrosymmetric matrix one is able to prove

THEOREM 9. (a) *If* $P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ *is an* $n \times n$ *centrosymmetric matrix with* $n = 2s$ *and* $A$, $B$, $C$, *and* $D$ $s \times s$ *matrices, then* $P$ *is orthogonally similar to*

$$Q^tPQ = \begin{pmatrix} A + JC & 0 \\ 0 & A - JC \end{pmatrix}.$$

(b) *If* $P = \begin{pmatrix} A & x & B \\ y & q & yJ \\ C & Jx & D \end{pmatrix}$ *is an* $n \times n$ *centrosymmetric matrix with* $n = 2s + 1$, $A$, $B$, $C$, *and* $D$ $s \times s$ *matrices,* $x$ *is a* $s \times 1$ *matrix,* $y$ *is a* $1 \times s$ *matrix, and* $q$ *is a scalar, then* $P$ *is orthogonally similar to*

$$Q^tPQ = \begin{pmatrix} A + JC & \sqrt{2}x & 0 \\ \sqrt{2}y & q & 0 \\ 0 & 0 & A - JC \end{pmatrix}.$$

*Proof*. The proof is given for part (a) and may also be found in [**4**]. Since $P = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is centrosymmetric,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix} = \begin{pmatrix} 0 & J \\ J & 0 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Hence $BJ = JC$ and $DJ = JA$. It then follows that $B = JCJ$ and $D = JAJ$. Thus $P = \begin{pmatrix} A & JCJ \\ C & JAJ \end{pmatrix}$ and multiplication gives

$$Q^tPQ = \begin{pmatrix} A + JC & 0 \\ 0 & A - JC \end{pmatrix}.$$

For part (b) one shows that $P = \begin{pmatrix} A & x & JCJ \\ y & q & yJ \\ C & Jx & JAJ \end{pmatrix}$ and multiplication gives

$$Q^tPQ = \begin{pmatrix} A + JC & \sqrt{2}x & 0 \\ \sqrt{2}y & q & 0 \\ 0 & 0 & A - JC \end{pmatrix}.$$

Theorem 9 is demonstrated for the $3 \times 3$ and $4 \times 4$ cases.

$$\begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & d \\ c & b & a \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} a + c & \sqrt{2}b & 0 \\ \sqrt{2}d & e & 0 \\ 0 & 0 & a - c \end{pmatrix},$$

$$\begin{pmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \end{pmatrix}\begin{pmatrix} a & b & c & d \\ e & f & g & h \\ h & g & f & e \\ d & c & b & a \end{pmatrix}\begin{pmatrix} 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \end{pmatrix}$$

$$= \begin{pmatrix} a + d & b + c & 0 & 0 \\ e + h & f + g & 0 & 0 \\ 0 & 0 & f - g & e - h \\ 0 & 0 & b - c & a - d \end{pmatrix}.$$

Using the partitioned matrix $Q^tPQ$ in the respective cases when $P$ is either of even or odd order, A. C. Aitken [**1**] shows that in the even case, using the Laplacian expansion, that

$$|P| = |Q^tPQ| = |A + JC||A - JC|.$$

In the odd case,

$$|P| = |Q^tPQ| = \begin{vmatrix} A + JC & \sqrt{2}\,x \\ \sqrt{2}\,y & q \end{vmatrix} |A - JC|.$$

In [**5**] I. J. Good shows in the even case that if $P = \begin{pmatrix} A & JCJ \\ C & JAJ \end{pmatrix}$ is a nonsingular matrix, then since $0 \neq |P| = |A + JC||A - JC|$, $A + JC$ and $A - JC$ are nonsingular matrices, and

$$P^{-1} = Q^t \begin{pmatrix} (A + JC)^{-1} & 0 \\ 0 & (A - JC)^{-1} \end{pmatrix} Q.$$

In the odd case, if $P$ is nonsingular, then $0 \neq |P| = \begin{vmatrix} A + JC & \sqrt{2}\,x \\ \sqrt{2}\,y & q \end{vmatrix} |A - JC|$ and $\begin{pmatrix} A + JC & \sqrt{2}\,x \\ \sqrt{2}\,y & q \end{pmatrix}$ and $(A - JC)$ are nonsingular matrices. Therefore

$$P^{-1} = Q^t \begin{pmatrix} \begin{pmatrix} A + JC & \sqrt{2}\,x \\ \sqrt{2}\,y & q \end{pmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & (A - JC)^{-1} \end{pmatrix} Q.$$

The formulas given in [**5**] by I. J. Good are more detailed and make the computation of the inverse of a centrosymmetric matrix a lot easier.

**5. Eigenvalues and Eigenvectors of a Centrosymmetric Matrix.** Using the matrix $Q$ described in Definition 8 of this paper, and observing that the $n \times n$ matrix $P$ is orthogonally similar to $\begin{pmatrix} A + JC & 0 \\ 0 & A - JC \end{pmatrix}$ in the even case and $\begin{pmatrix} A + JC & \sqrt{2}x & 0 \\ \sqrt{2}y & q & 0 \\ 0 & 0 & A - JC \end{pmatrix}$ in the odd case, one can easily see that the eigenvalues of $P$ are the same as those of $A + JC$ and $A - JC$ in the even case and the same as those of $\begin{pmatrix} A + JC & \sqrt{2}x \\ \sqrt{2}y & q \end{pmatrix}$ and $A - JC$ in the odd case. It needs to be noted at this time that I. J. Good goes into more detail in describing the eigenvalues of a centrosymmetric matrix on p. 928 of [**5**].

After the eigenvalues of a matrix have been found, one is confronted with the problem of finding the eigenvectors. In [**2**] A. L. Andrew presents the following definition. The two results that follow may be found in [**2**] and [**4**].

DEFINITION 10. An $n$ dimensional vector $X$ is symmetric if $JX = aX$ with $a = 1$ and is skew-symmetric if $a = -1$.

Using Definition 10 one easily obtains

THEOREM 11. *If $P$ is an $n \times n$ centrosymmetric matrix, then the eigenvectors corresponding to the eigenvalue $\lambda$, where* $\dim NS(\lambda I - P) = 1$, *are either symmetric or skew-symmetric.*

*Proof.* If $X$ is an eigenvector of $P$ corresponding to $\lambda$, then $PX = \lambda X$ and $PJX = \lambda JX$. It then follows that $aX = JX$ for some scalar $a$ not equal to 0, and $a$ is also an eigenvalue of $J$. Since $J$ is an orthogonal matrix, its eigenvalues are $\pm 1$, and so $a = \pm 1$. Therefore $JX = \pm X$, and $X$ is either symmetric or skew-symmetric.

A corollary of Theorem 11 which is useful in computing eigenvectors of a centrosymmetric matrix for a given eigenvalue $\lambda$, when $\dim NS(\lambda I - P) > 1$, is given below.

COROLLARY 12. *If $P$ is a centrosymmetric matrix and $X$ is an eigenvector of $P$ corresponding to*

*the eigenvalue* $\lambda$, *then* $JX$ *is also an eigenvector of* $P$ *corresponding to* $\lambda$.

*Proof*. If $PX = \lambda X$, then $PJX = \lambda JX$ and $JX$ is an eigenvector of $P$ corresponding to the eigenvalue $\lambda$.

**6. Examples.** If $P = \begin{pmatrix} 1 & 0 & 0 \\ a & b & a \\ 0 & 0 & 1 \end{pmatrix}$, where the row sums are equal, then $(0,1,2)^t$ is an eigenvector of $P$ corresponding to the eigenvalue 1. From Corollary 12 it is easy to see that $(2,1,0)^t$ is also an eigenvector of $P$ corresponding to the eigenvalue 1.

In papers **[3]** and **[14]** explicit formulas are given for the eigenvalues and eigenvectors of two classes of centrosymmetric matrices. It is shown in **[3]** that $\lambda_t = n_t/n^t$, for $0 \leqslant t \leqslant n$, are the eigenvalues of the centrosymmetric matrix $P$, where

$$P_{i+1,j+1} = \binom{n}{j}\left(\frac{n-i}{n}\right)^{n-j}\left(\frac{i}{n}\right)^{j}, \quad 0 \leqslant i,\, j \leqslant n.$$

This information may also be found in **[7]**, but Theorem 2 of **[3]** gives an explicit formula for the eigenvectors of $P$. For $n = 8$ and $t = 7$, $\lambda_7 = 8_7/8^7 = (8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2)/8^7$ is an eigenvalue and each $(i+1)$th entry of an eigenvector $X_7$ is any convenient multiple of the determinantal expression

$$X_{i,7} = \left(\frac{1}{2}n - 1\right)(n-i)i \begin{vmatrix} 1 & 1-(n-3)_4 n^{-4} & 0 \\ (n-i)i & n-1 & 1-(n-5)_2 n^{-2} \\ (n-i)^2 i^2 & (n-1)^2 & 3n-5 \end{vmatrix}.$$

The first 6 entries of $X_7$ are computed and, after suppressing the factor $315/2^{14}$, are found to be 0, 2835, $-7552$, 7705, 0, and $-7705$. Since $\dim NS(P - \lambda_7 I) = 1$, $-7705 = a(7705)$ and $a = -1$. From Theorem 11 it then follows that 7552, $-2835$, and 0 are the last three entries of $X_7$.

In **[14]** a centrosymmetric matrix $P$ is given by

$$P_{i+1,j+1} = \binom{(n-i)\alpha}{(n-j)}\binom{i\alpha}{j}\Big/\binom{n\alpha}{n}, \quad 0 \leqslant i,\, j \leqslant n.$$

The eigenvalues of this matrix are given by $\lambda_t = (\alpha^t n_t)/(\alpha n)_t$ for $0 \leqslant t \leqslant n$ and $\alpha$ a positive integer greater than 1. Theorem 2 of **[14]** gives an explicit formula for the entries of the eigenvectors for each eigenvalue. In the case $n = 7$ and $t = 4$ an eigenvalue is $\lambda_4 = (\alpha^4 7_4)/(\alpha 7)_4$ and the first 5 terms of an eigenvector $X_4$ corresponding to $\lambda_4$ are found to be 0, $90(\alpha - \frac{1}{6})$, $5(\alpha + 1)$, $-81(\alpha - 8/27)$, and $-81(\alpha - 8/27)$. Since $X_4 = aX_4^R$, $a = 1$, and the last three terms are $5(\alpha + 1)$, $90(\alpha - \frac{1}{6})$, and 0.

**7. Conclusion.** In conclusion it is worthwhile to note that if $\phi$ is the mapping defined by $\phi$: $P \to Q^t P Q$, where $P$ is a centrosymmetric matrix and $Q$ is the orthogonal transformation defined in Definition 8, then $\phi$ is an isomorphism from the algebra of centrosymmetric matrices to the direct sum of two full matrix algebras.

THEOREM 13. *The algebra of all real centrosymmetric matrices of order n is isomorphic to the direct sum* $M_{n-k}(R) \oplus M_k(R)$ *(where* $k = [n/2]$*) of full matrix algebras.*

**References**

1. A. C. Aitken, Determinants and Matrices, Oliver and Boyd, Edinburgh, 1939.

2. A. L. Andrew, Eigenvectors of certain matrices, Linear Algebra Appl., 7 (1973) 157–162.

3. F. Barnett, J. S. Frame, and J. Weaver, Eigenvalues and eigenvectors of a certain stochastic matrix, Linear and Multilinear Algebra, 13 (1983) 345–350.

4. A. Cantoni and P. Butler, Eigenvalues and eigenvectors of symmetric centrosymmetric matrices, Linear Algebra Appl., 13 (1976) 275–288.

5. I. J. Good, The Inverse of a centrosymmetric matrix, Technometrics, 12 (1970) 925–928.
6. F. Graybill, Introduction to Matrices with Applications in Statistics, Wadsworth, Belmont, CA, 1969.
7. M. Iosifescu, Finite Markov Processes and Their Applications, Wiley, New York, 1980.
8. R. G. Khazanie, An indication of the asymptotic nature of the Mendelian Markov process, J. Appl. Probability, 5 (1968) 350–356.
9. R. G. Khazanie and H. E. McKean, A Mendelian Markov process with binomial transition probabilities, Biometrika, 53 (1966) 37–48.
10. M. Kimura, Some problems of stochastic processes in genetics, Ann. of Math. Stat., 28 (1957) 882–901.
11. G. Malécot, Sur un problème de probabilité en chaîne que pose la génétique, C. R. Acad. Sci., Paris, 219 (1944) 279–281.
12. A. A. Markov, Extension of the law of large numbers to dependent variables, Izv. Fiz.-Mat. Obšč, pri. Kazansk. Univ. (2 Ser.) 15 (1906) 135–156.
13. I. Schensted, Model of subnuclear segregation in the macronucleus of ciliates, The American Naturalist, XCII (1958) 161–170.
14. J. Weaver, Eigenvalues and eigenvectors of a certain stochastic matrix II, Linear and Multilinear Algebra, 13 (1983) 351–355.
15. S. Wright, Evolution in Mendelian populations, Genetics, 16 (1931) 97–159.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## MONTHLY UNSOLVED PROBLEMS, 1969–1985

RICHARD K. GUY
*Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4*

As usual in these updating articles in odd-numbered years, references in brackets [1983, 684] are year and page numbers of this MONTHLY, while references in parentheses (1981) are to the bibliography at the end; items not yet published are labelled (tbp) or (wrc), according as they are likely to be published formally or to remain as written communications.

In order to enhance the usefulness of the bibliography, we have added Math. Reviews numbers and a bracket indicating where the problem originally appeared in the MONTHLY.

Carmichael's problem, find $n$ such that $\varphi(x) = n$ has a unique solution, $x$, where $\varphi(x)$ is Euler's totient function, was mentioned by Klee [1969, 288]. The unitary analog (a unitary divisor, $d$, of $n$, is one where $d$ and $n/d$ are coprime) was discussed by Ismail and Subbarao (1976), a reference we didn't notice earlier.

Graceful numbering of trees and other graphs, an offshoot of the Kotzig-Ringel problem asked by Duke [1969, 1128] continues to generate papers. Label the $n + 1$ vertices of any tree with the integers $0, 1, \ldots, n$ so that the differences along the $n$ edges are $1, 2, \ldots, n$. Prospective authors are urged to consult the existing literature. My bibliography now contains 109 items; recent additions are Poljak and Sura (1982), Bloom and Hsu (1983) and Grace (1982, 1983).

Ogilvy [1970, 388; 1975, 996; 1983, 684] asked about the convergence of $z_{n+1} = \sqrt{c + z_n}$ and partial answers were given by Gerber and by Rohde. Peter Walker (1983) notes theorems implicit in Burckel (1981) who quotes Macintyre (1966):

Theorem 1: Let $f$ be an analytic map of the disc $D = \{z: |z| < 1\}$ to itself which is not a rotation, and with $f(0) = 0$, then for any $z_0$ in $D$ and $z_{n+1} = f(z_n)$, $n \geqslant 0$, we have $z_n \to 0$ as $n \to \infty$. Theorem 2: Let $f$ be an analytic mapping of a simply connected region $G \subset C$ to itself with $f(G) \neq G$, and suppose that $f$ has a fixed point at $l$ in $G$, then for any $z_0$ in $G$, the sequence defined by $z_{n+1} = f(z_n)$ for $n \geqslant 0$ converges to $l$ as $n \to \infty$.

We can now give the reference to Goodey's (1984) confirmation of Peterson's [1972, 505; 1983, 684] conjectures. If $S, C$ are non-coincident planar convex sets, then they have equal width in every direction just if $\alpha(S, C')$, the number of connected components of $\partial S \cap \partial C'$, is even or infinite for all translations $C'$ of $C$.

In confirmation of a conjecture of Erdős and Guy [1973, 52], Leighton (1981) proves that the crossing number of a graph with $n$ vertices and $k \geqslant 4n$ edges is at least $k^3/375n^2$. A similar theorem was reported [1981, 757] to be in preprints which were later superseded by Ajtai, Chvátal, Newborn and Szemerédi (1982) who prove "at least $k^3/100n^2$ edge-crossings and fewer than $10^{13n}$ crossing-free subgraphs." The latter result settles a conjecture of Newborn and Moser (1980).

Detlef Seese's results on the computability of the crossing number and rectilinear crossing number of a graph have also been obtained by Garey and Johnson (1983), who also prove that the problem of calculating the crossing number of a graph is NP-complete. Seese uses this result to show that the rectilinear crossing problem is also NP-complete. He mentioned this at ICM, Warsaw, but doesn't yet plan formal publication.

Jan van Leeuwen sends a useful short history of Nash's reachability problem [1973, 292]. He himself (1974) solved the problem for dimension $\leqslant 3$, and Hopcroft and Pansiot (1976) for dimension $\leqslant 5$. The complete problem, for every dimension, was settled by Ernst Mayr (1980, 1981, 1984); and see Mayr and Meyer (1981).

We indicated [1979, 848] that our earlier announcement [1977, 808–809] of the solution was premature. Grabowski (1980) obtained "a positive solution to a decision problem left open by Karp and Miller (1969) and Hack (1976) and posed explicitly by Landweber and Robertson (1978): the problem of persistence of a vector addition system (or a Petri net)." Incidentally it gives a positive solution to the reachability problem for persistent vector addition systems. See also Kleine-Büning (1980). Kosaraju (1982) has provided a somewhat more elementary proof.

A famous theorem of Vizing is that if the maximum valence of a graph is $\rho$, then the edge-chromatic number is $\rho$ or $\rho + 1$. A graph of edge-chromatic number $\rho + 1$ is called **critical** if deletion of any edge reduces its edge-chromatic number. Wilson and Beineke [1976, 128] conjectured that there are no critical graphs with an even number of vertices. An infinity of counterexamples was found by Goldberg in 1978; see Goldberg (1981) or Chetwynd and Wilson (1983).

Alter [1980, 43] revived the thousand-year-old congruent number problem: which integers $a$ allow simultaneous solution of the diophantine equations

$$x^2 + ay^2 = z^2 \quad \text{and} \quad x^2 - ay^2 = t^2.$$

He conjectured that if $a \equiv 5$, 6 or 7 (mod 8), then $a$ is congruent in this sense. We reported [1983, 685] that Tunnell had proved that all sixteen numbers $\equiv 1$, 2, or 3 (mod 8) for which a blank was left in Table 1 [1981, 759] or in Table 7 of Guy (1981), were non-congruent, but we did not give the reference (1983).

Jean Lagrange, in a letter dated 83:02:10, reported that the nine remaining numbers $\equiv 5$ or 7 (mod 8), of unknown status in the same table, are now known to be congruent, so the conjecture is verified for $a < 1000$.

Gerd Hofmeister notes that it would have been more appropriate, in my comments [1981, 759] on the postage-stamp problem [1980, 206] to have quoted Rødseth's papers (1981, 1982) on $h$-bases, rather than his papers on the Frobenius problem.

Classification of dendrograms is of interest to theoretical biologists, so Smith and Waterman asked [1980, 552]: how alike are two trees? They quoted their own paper (1978), but Jarvis,

Luedeman and Shier (1983, 1984) have since found counterexamples to Theorem 3. These also invalidate the proofs of Theorems 4 and 5: neither rigorous proofs nor counterexamples to these theorems have been found. See also Boland, Brown and Day (1983), Brown (1982), Čulík and Wood (1983), Day (1983), Robinson (1971) and Moore, Goodman and Barnabas (1973).

Emma Lehmer has kindly written us the definitive note about the next problem.

## ON SQUARE-SEPARABLE PRIMES

EMMA LEHMER
*1180 Miller Avenue, Berkeley, CA 94708*

Jacob Beard [1980, 744] defined a 3-**square-separable prime** as a prime such that each pair of cubic residues which are not quadratic residues, is separated by a quadratic residue. Richard Guy [1980, 686] stated "it seems likely that 7, 19, 37 and 43 are the only 3-square-separable primes," and asked the corresponding question where 3 is replaced by any odd $s$.

It is the purpose of this short note to restate this question in terms of cyclotomy of order $2s$, to answer the question for $s = 3$ and $s = 5$ and to indicate how it can be done for larger values of $s$.

Let $p = 2sf + 1$ be a prime with $s$ odd. We partition the integers less than $p$ into $2s$ sets $C_j$ with respect to a primitive root $g$, so that

$$C_j = \left\{ g^{2sk+j} \,(\mathrm{mod}\ p), 0 \leqslant k \leqslant f-1 \right\}, \qquad 1 \leqslant j \leqslant 2s.$$

The cyclotomic number $(i, j)_{2s}$ is the number of integers in $C_i$ which are followed by an integer of $C_j$. For example, if $p = 61$, $s = 3$, $f = 10$, one of these sixth power residue classes is

$$C_3 = \{8, 24, 11, 33, 38, 53, 37, 50, 28, 23\}.$$

It contains the two consecutive pairs 23, 24 and 37, 38, so that the cyclotomic number $(3,3)_6$ is 2 for this prime.

In terms of this notation we can state at once the following

THEOREM 1. *The prime $p = 2sf + 1$ is not s-square-separable if $(s, s)_{2s} > 0$.*

*Proof.* $(s, s)_{2s} > 0$ implies that there is at least one pair of consecutive $s$-power residues.

It is well known that there is only a finite number of primes for which $(i, j)_{2s} = 0$. Hence our problem reduces to finding and examining those primes for which $(s, s)_{2s} = 0$.

Another obvious theorem is:

THEOREM 2. *If $p = 8n + 5$ and if 2 is an s-power residue, then $p = 2sf + 1$ is not s-square separable.*

*Proof.* Under the conditions of the theorem, the middle pair $(p-1)/2, (p+1)/2$ are both $s$-power residues and quadratic non-residues, and hence $p$ is non-separable.

Previous authors considered $p = 7$ to be 3-square separable and $p = 11$ to be 5-square-separable, but since these primes, and in general the prime $p = 2s + 1$, have only a single $s$-power residue, such primes may be omitted.

For $s = 3$ the cyclotomic numbers of order six can be expressed in terms of the quadratic partition [**2**]

$$4p = L^2 + 27M^2 \text{ with } L \equiv 1 \,(\mathrm{mod}\ 3), \text{ and, if } LM \text{ is odd}, \quad M \equiv -L \,(\mathrm{mod}\ 4).$$

In particular, the value of $(3,3)_6$ is given by

$$36(3,3)_6 = \begin{cases} p - 5 - 2L & f \text{ even, 2 a cubic residue} \\ p - 11 + 4L & f \text{ odd, 2 a cubic residue} \\ p - 5 + (5L + 9M)/2 & f \text{ even, 2 a cubic non-residue} \\ p - 11 - (L + 9M)/2 & f \text{ odd, 2 a cubic non-residue.} \end{cases}$$

Since $L < \sqrt{4p}$ and $M < \sqrt{4p/27}$, it follows that $(3,3)_6 = 0$ implies that $p < 83$ in all cases. The following table gives the values of $L, M$ and $(3,3)_6$ for $p < 83$:

| $p$ | 7 | 13 | 19 | 31 | 37 | 43 | 61 | 67 | 73 | 79 |
|---|---|---|---|---|---|---|---|---|---|---|
| $L$ | 1 | −5 | 7 | 4 | −11 | −8 | 1 | −5 | 7 | −17 |
| $M$ | −1 | 1 | 1 | 2 | −1 | 2 | 3 | −3 | −3 | 1 |
| $(3,3)_6$ | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 2 | 2 |

The primes 13, 19, 37 and 43 have to be checked for square-separability, but this is easily done with a table of indices. This proves that 13, 37 and 43 are indeed the only 3-square-separable primes.

Similarly for $s = 5$ we can write down the values of $(5,5)_{10}$ in terms of the quadratic partition [4]

$$16p = x^2 + 50u^2 + 50v^2 + 125w^2, \quad x \equiv 1 \pmod 5)$$

as follows:

$$100(5,5)_{10} = \begin{cases} p - 9 + 2x & f \text{ even, 2 a quintic residue} \\ p - 19 + 2x & f \text{ odd, 2 a quintic residue} \\ p - 9 + (17x + 50u - 25w)/4 & f \text{ even, 2 a quintic non-residue} \\ p - 19 + (7x - 50u + 25w)/4 & f \text{ odd, 2 a quintic non-residue.} \end{cases}$$

Since $x < 4\sqrt{p}$, $u, v < 4\sqrt{p/50}$, $w < 4\sqrt{p/125}$, it follows that $(5,5)_{10} = 0$ implies that $p \leqslant 911$. If we calculate $(5,5)_{10}$ from the above formulas, we find that the only values for which $(5,5)_{10} = 0$ are $p = 11, 31, 41, 61, 71$, and 101. When we examine these we find that 31, 41 and 71 are indeed 5-square-separable, while 61 is not since it does not contain a square between the pair 29 and 32 of quintic residues; similarly $p = 101$ has a pair 39 and 41 of quintic residues with the quadratic non-residue 40 in between. Unfortunately, Guy listed 61 and 101 as 5-square-separable primes.

We have proved:

THEOREM 3. *The only* 3-*square-separable primes are* $p = 13, 37$ *and* 43. *The only* 5-*square-separable primes are* $p = 31, 41$ *and* 71.

Cyclotomy of order 14 has been developed by Muskat [3] and of order 18 by Baumert and Fredricksen [1]. These could be used to make the corresponding analysis of the problem for $s = 7$ and $s = 9$, but this would involve more cases and probably should be computer assisted.

## References

1. Leonard D. Baumert and H. Fredricksen, The cyclotomic numbers of order eighteen with applications to differences sets, Math. Comp., 21 (1967) 204–219; T20-35 = MR 36 #6370.

2. D. H. and Emma Lehmer, The sextic period polynomial, Pacific J. Math., 111 (1984) 341–355.

3. Joseph B. Muskat, The cyclotomic numbers of order fourteen, Acta Arith., 11 (1965/66) 263–279; T20-29 = MR 33# 1302.

4. Albert Leon Whiteman, The cyclotomic numbers of order ten, Proc. Sympos. Appl. Math., 10 (1960) 95–111; T20-24 = MR 22# 4862.

Göbel's rind-peeling problem [1982, 113; 1983, 686] remains baffling in its simplicity. Schwenk (tbp, 1985) considered sequences containing just two copies of each of $m$ letters and used a computer to find sequences yielding the maximum number, $\rho(m)$, of rinds. Had all $2m$ letters been distinct, the number would be $2^{2m-1}$. The **deficit**, $d = 2^{2m-1} - \rho(m)$ for the best of these sequences is, for $m \leqslant 14$:

| $m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d$ | 1 | 2 | 8 | 18 | 55 | 138 | 470 | 1164 | 4055 | 10140 | 35609 | 89782 | 316513 | 803040 |

Schwenk conjectures that, for $m \geqslant 8$, the greatest number of rinds is given by the sequence $T(k)$, $k + T(m - k)$, where $k = \lfloor m/2 \rfloor$ and $T(k)$ is the sequence $k, k-1, \ldots, 4, 2, 3, 1, 1, 2, 3, 4, \ldots, k$, which runs from $k$ to 1 and back to $k$, except for the reversal of 3, 2. Gibson and Slater (1982, tbp) also consider this case, and generalize the problem to peeling the labelled (or colored) vertices from a graph, subject to the condition that the graph remains connected.

In spite of editorial admonition [1983, 35] and the appearance of Jeff Lagarias's article [1985, 3–23], the infamous $3x + 1$ problem maintains its lure. Korec & Znám (wrc) let $Y$ be a subset of $P$, the set of all positive integers, and write $P \prec Y$ if, for every $x$ in $P$, there is some $y$ in $Y$ and $i$ and $j$ such that $f^i(y) = f^j(x)$, where $f(x) = x/2$ if $x$ is even and $3x + 1$ otherwise. The $3x + 1$ problem then becomes:

$$¿P \prec \{1\}?$$

They prove that if 2 is a primitive root of $p^2$, then for every $a, n$ in $P$, where $p$ does not divide $a$,

$$P \prec \{a, \bmod p^n\}.$$

They say that they have also proved that $P \prec \{1, \bmod 4\}$ and that $P \prec X$ for a set $X$ of zero density.

Somewhat to his surprise, Peter Giblin writes that moving triangles of the kind envisioned [1983, 121] do exist; see his paper with Kingston (tbp). This was also discovered by David Lindsay (wrc).

Leonard Gilman extends the age of snowplow problems [1983, 199] to at least 43 years: such a problem appeared in the *first* edition (1942) of Agnew's book on Differential Equations.

Jan Mycielski writes concerning his pursuit problem [1984, 415] that fact (iii) implies that $T_0(v, x)$ is not differentiable at $x = 1 - v$. However, problem (c): Is $T_0(v, x)$ differentiable? is still open for all other $x \in [0, 1]$. His paper [1984, 416, **4**] is available on request.

Thews [1984, 416] asked if the circle $|z| < rv/(v - 1)$ was the domain of largest area which could be guarded by a defender with destruction radius $r$ and maximum speed 1 against an invader with maximum speed $v > 1$. Rzymowski and Stachura (wrc) show that this is so.

Johnson and Brooks (wrc) succeed in generating a bivariate Poisson distribution [1984, 562] with correlation $\rho$, for any $\rho$, $-\ln 2 < \rho < 0$. Let $U$ be distributed uniformly on [0, 1], $X_1 = -\ln U$, $X_2 = -\ln(1 - U)$. Then $X_1, X_2$ are exponentially distributed with parameter $\lambda = 1$. Define $N_i(t)$, $i = 1, 2$, as a counting process with interarrival times according to $X_i$. Then the vector $(N_1(t), N_2(t))$ is bivariate Poisson. For intervals of length $t < \ln 2$,

$$\Pr(N_1(t) > 0,\, N_2(t) > 0) = 0, \qquad \Pr(N_1(t) = 0,\, N_2(t) = 0) = 2e^{-t} - 1 > 0$$

and

$$\Pr(N_1(t) = k,\, N_2(t) = 0) = \Pr(N_1(t) = 0,\, N_2(t) = k) = t^k e^{-k}/k!, \qquad k = 1, 2, \ldots.$$

Further,

$$E(N_1(t)) = E(N_2(t)) = 0, \qquad \mathrm{VAR}(N_1(t)) = \mathrm{VAR}(N_2(t)) = t$$

and correlation $(N_1(t), N_2(t)) = -t$.

Roger Nelsen (wrc) has also generated such distributions with correlation coefficients in the interval $-0.5 \geqslant \rho \geqslant -0.9585$.

Thiemann [1984, 562] asked: when is there a probability measure on a $\sigma$-algebra? Grzegorz Plebanek (wrc) of Wrocław sends the following result: Let $A$ be a set and let $\mathscr{A} = \{A_\alpha : \alpha \in \Delta\}$ be a **semi-compact** family of subsets of $A$, i.e., for each subset $\mathscr{A}'$ of $\mathscr{A}$ with $\cap\mathscr{A}' = \varnothing$, there is a finite subset $\mathscr{A}''$ of $\mathscr{A}'$ such that $\cap\mathscr{A}'' = \varnothing$. $\{a_\alpha : \alpha \in \Delta\}$ is any set of real numbers. By $a(\mathscr{A})$ [resp. $\sigma(\mathscr{A})$] we denote an algebra [resp. $\sigma$-algebra] generated by the family $\mathscr{A}$. If, for every finite subset $S$ of $\Delta$, there exists a quasi-measure $\mu_S$ on $a(\mathscr{A})$ such that $\mu_S(A_\alpha) \geqslant a_\alpha$ for $\alpha \in S$, then there exists a measure $\mu$ on $\sigma(\mathscr{A})$ satisfying $\mu(A_\alpha) \geqslant a_\alpha$ for every $\alpha$. He proves this from results of Pachl (1978, lemma 3.3), Topsoe (1974) and a classical result of Marczewski (1953).

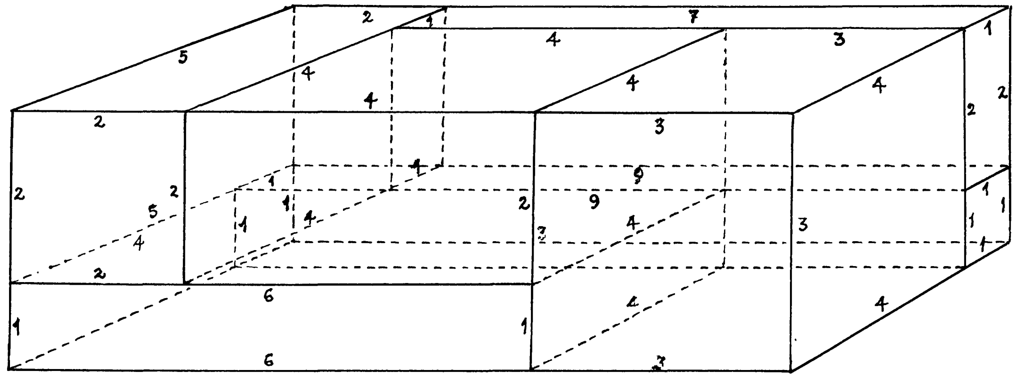


FIG. 1

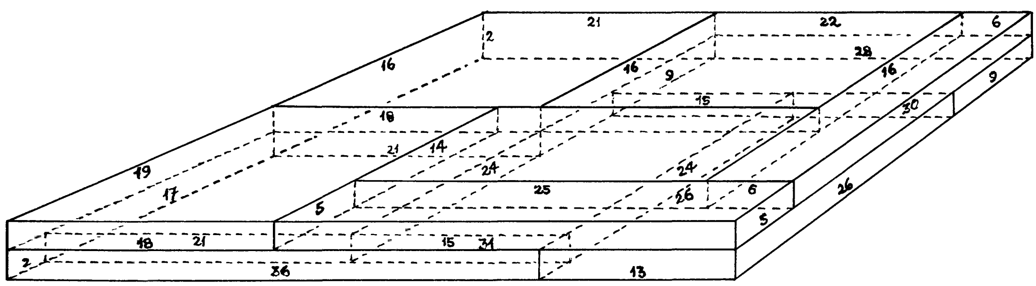


FIG. 2

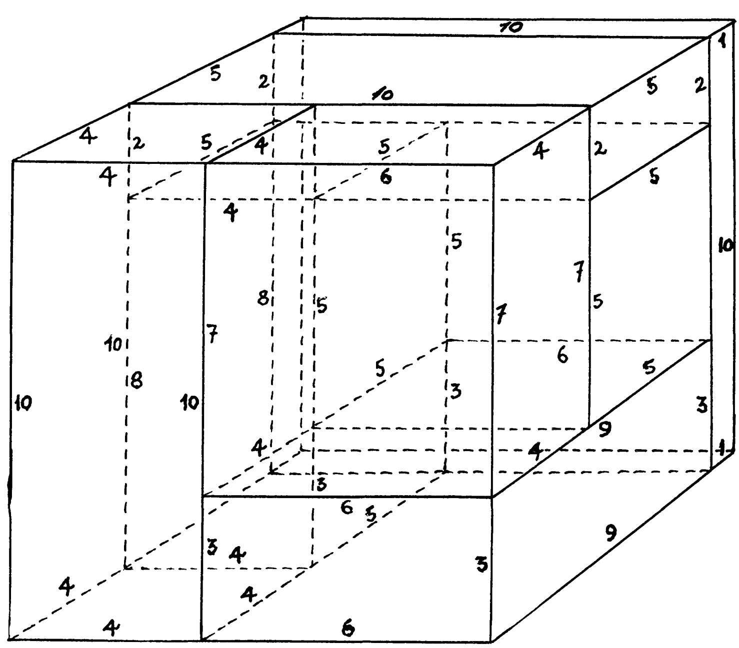


FIG. 3

Charles H. Jepsen (tbp) discovered six incomparable cuboids [1984, 624] which will tile a $3 \times 5 \times 9$ cuboid (Fig. 1).

He later proved that this example,

$$(1 \times 1 \times 9) + (1 \times 2 \times 7) + (1 \times 4 \times 6) + (2 \times 2 \times 5) + (2 \times 4 \times 4) + (3 \times 3 \times 4) = 135,$$

has minimum volume. He proposes two further problems:

If we erect prisms of height 2 on the seven pieces that tile a $13 \times 22$ rectangle [1984, 625, Fig. 2] we have a trivial incomparable tiling of a $2 \times 13 \times 22$ cuboid. What is the $2 \times B \times C$ cuboid of least volume that admits a non-trivial tiling (one in which the pieces don't all share a common dimension)? Fig. 2 illustrates the smallest found by Jepsen. It derives from two tilings of a $35 \times 49$ rectangle, having a $16 \times 21$ rectangle in common. One piece is $2 \times 16 \times 21$. The other ten all have thickness 1, and other dimensions $18 \times 19$, $17 \times 21$, $16 \times 22$, $15 \times 24$, $14 \times 25$, $13 \times 26$, $9 \times 28$, $6 \times 30$, $5 \times 31$ and $2 \times 36$: the 1st, 3rd, 5th, 8th and 9th are in the top layer.

Alternatively, is it possible to tile a $2 \times B \times C$ cuboid with incomparable $1 \times p \times q$ pieces? I.e., are there two tilings of the same rectangle, with $m$ and $n$ pieces, say, and all $m + n$ pieces incomparable?

Later still, Jepsen sends the non-trivial incomparable tiling of a cube shown in Fig. 3:

$$(1 \times 10 \times 10) + (2 \times 5 \times 10) + (3 \times 6 \times 9) + (4 \times 4 \times 10) + (4 \times 5 \times 8) + (4 \times 6 \times 7) + (5 \times 5 \times 6) = 10^3.$$

He can prove that the number of pieces, seven, is minimal, and he believes that the edge-length, 10, is minimal, too.

My indebtedness to innumerable correspondents is clear. Thanks, too, to several readers who said that they find the section enjoyable as well as instructive. I will make a New Year resolution that it will appear more regularly.

## Bibliography


M. Ajtai, V. Chvátal, M. M. Newborn, and E. Szemerédi, Crossing-free subgraphs, Ann. Discrete Math., 12 (1982) 9–12 [1973, 52].

Gary S. Bloom and D. Frank Hsu, Graceful directed graphs, Report #CCNY-CS283, Comput. Sci., City Coll. New York, 1983 [1969, 1128].

Ralph P. Boland, Edward K. Brown, and William H. E. Day, Approximating minimum-length-sequence metrics: a cautionary note, Math. Social Sci., 4 (1983) 261–270; MR 84m:92042 [1980, 552].

Edward K. Brown, Some closest partition distances between trees, B.Sc. Hons. Thesis, Comput. Sci., Memorial Univ. Newfoundland, May 1982 [1980, 552].

R. B. Burckel, Iterating analytic self-maps of discs, this MONTHLY, 88 (1981) 396–407; MR 82g:30046 [1969, 1128].

Amanda G. Chetwynd and Robin J. Wilson, The rise and fall of the critical graph conjecture, J. Graph Theory, 7 (1983) 153–157; MR 84h:05042 [1976, 128].

Karel Čulík II and Derick Wood, A note on some tree similarity measures, Inform. Process. Lett., 15 (1982) 39–42; MR 83k:68059 [1980, 552].

William H. E. Day, Properties of the nearest neighbor interchange metric for trees of small size, J. Theoret. Biol., 101 (1983) 275–288; MR 84i:92083 [1980, 552].

M. R. Garey and D. S. Johnson, Crossing-number is NP-complete, SIAM J. Algebraic Discrete Methods, 4 (1983) 312–316; MR 85a:68070 [1973, 52].

Peter J. Giblin and J. G. Kingston, Caustics by reflexion in the plane with stable triple crossings (tbp) [1983, 121].

Peter M. Gibson and Peter J. Slater, The rinding number of a graph, *82T-05-322, Abstracts Amer. Math Soc., 3 (1982) 285–286 [1982, 113].

Peter M. Gibson and Peter J. Slater, Rinds of a graph, Ars Combin. (tbp) [1982, 113].

Mark K. Goldberg, Construction of class 2 graphs with maximum vertex degree 3, J. Combin. Theory Ser. B, 31 (1981) 282–291; MR 84c:05030 [1976, 128].

P. R. Goodey, Intersections of planar convex curves. Math. Ann., 267 (1984) 145–159 [1972, 505].

Jan Grabowski, The decidability of persistence for vector addition systems, Inform. Process. Lett., 11 (1980) 20–23; MR 82c:68035 [1973, 292].

Thom Grace, Graceful, harmonious and sequential graphs, Ph.D. thesis, Univ. Illinois Chicago Circle, 1982 [1969, 1128].

Thom Grace, On sequential labellings of graphs, J. Graph Theory, 7 (1983) 195–201; MR 84m:05075 [1969, 1128].

Richard K. Guy, Problem D27, Unsolved Problems in Number Theory, Springer, 1981 [1980, 43].

M. Hack, Decidability questions for Petri nets, Report TR-161, MIT Lab. Comput. Sci., 1976 [1973, 292].

M. Hack, The equality problem for vector addition systems is undecidable, Theoret. Comput. Sci., 2 (1976) 77–95; MR 55#12496 [1973, 292].

John Hopcroft and Jean-Jacques Pansiot, On the reachability problem for 5-dimensional vector addition systems, TR 76-280, Dept. Comput. Sci., Cornell Univ., New York, 1976, Theoret. Comput. Sci., 8 (1979) 135–159; MR 80h:68022 [1973, 292].

Mourad Ismail and M. V. Subbarao, Unitary analogue of Carmichael's problem, Indian J. Math., 18 (1976) 49–55; MR 82d:10011 [1969, 288].

J. P. Jarvis, John K. Luedeman, and Douglas R. Shier, Comments on computing the similarity of binary trees, J. Theor. Biol., 100 (1983) 427–433; MR 84c:92003 [1980, 552].

J. P. Jarvis, John K. Luedeman, and Douglas R. Shier, Counterexamples in measuring the distance between binary trees, Math. Sci., Clemson Univ., July 1982 [1980, 552].

Charles H. Jepsen, Tiling with incomparable cuboids, Math. Mag., 59 (1986) [1984, 624].

Mark E. Johnson and Daniel G. Brooks, Solution to "Can we generate a bivariate Poisson distribution with a negative correlation?" (wrc) [1984, 562].

I. Korec and Štefan Znám, A note to the $3x + 1$ problem (wrc) [1983, 35].

R. M. Karp and R. E. Miller, Parallel program schemata, J. Comput. System Sci., 3 (1969) 147–195; MR 39 #8024 [1973, 292].

Hans Kleine-Büning, Decision problems in generalized vector addition systems, Fund. Inform. (4) 3 (1980) 497–511; MR 83c:68034 [1973, 292].

S. Rao Kosaraju, Decidability of reachability in vector addition systems, Proc. 14th ACM Sympos. Theory Comput., San Francisco, 1982, 267–281 [1973, 292].

L. H. Landweber and E. L. Robertson, Properties of conflict-free and persistent Petri nets, J. Assoc. Comput. Mach., 25 (1978) 352–364; MR 58 #9913 [1973, 292].

J. van Leeuwen, A partial solution to the reachability problem for vector addition systems, Proc. 6th ACM Sympos. Theory Comput., Seattle, 1974, 303–309; MR 54#1725 [1973, 292].

Frank Thompson Leighton, Layouts for the shuffle-exchange graph and lower bound techniques for VLSI, Ph.D. thesis, MIT Lab. Comput. Sci., 81:08:07 and Report MIT/LCS/TR-274, June 1982, and Complexity Issues in VLSI, MIT Press, 1983, esp. pp. 83–85 [1973, 52].

A. J. Macintyre, Convergence of $i^i$, Proc. Amer. Math. Soc., 17 (1966) 67; MR 32 #5855 [1970, 388].

E. Marczewski, On compact measures, Fund. Math., 40 (1953) 113–124; MR 15-610 [1984, 562].

Ernst W. Mayr, Ein Algorithmus für das allgemeine Erreichbarkeits problem bei Petrinetzen und damit zusammenhängende Probleme, Ph.D. thesis, Report TUM-18010, Inst. Inform., Tech. Univ., München, 1980 [1973, 292].

Ernst W. Mayr, An algorithm for the general Petri-net reachability problem, Proc. 13th ACM Sympos. Theory Comput., Milwaukee, 1981, 238–246 [1973, 292].

Ernst W. Mayr, Persistence of vector replacement systems is decidable, Acta Inform., 15 (1981) 309–318; MR 82f: 68060 [1972, 292].

Ernst W. Mayr, An algorithm for the general Petri net reachability problem, SIAM J. Comput., 13 (1984) 441–460 [1973, 292].

Ernst W. Mayr and Albert R. Meyer, The complexity of the finite containment problem for Petri nets, J. Assoc. Comput. Mach., 28 (1981) 561–576; MR 82f:68061 [1973, 292].

G. William Moore, M. Goodman, and J. Barnabas, An iterative approach from the standpoint of the additive hypothesis to the dendrogram problem posed by molecular data sets, J. Theoret. Biol,. 38 (1973) 423–457 [1980, 562].

Roger B. Nelsen, Bivariate Poisson distributions with negative correlation, (wrc) [1984, 562].

M. Newborn and W. O. J. Moser, Optimal crossing-free Hamiltonian circuit drawings of $K_n$, J. Combin. Theory Ser. B, 29 (1980) 13–26; MR 81h:05088 [1973, 52].

J. Pachl, Disintegration and compact measures, Math. Scand., 43 (1978/79) 157–168; MR 80d:28020 [1984, 562].

Svatopluk Poljak and Miroslav Sura, On a construction of graceful trees, Proc. 3rd Czech. Sympos. Graph Theory, Prague, 1982, Teubner-Texte für Math., 59 (1983) 220–222 [1969, 1128].

Svatopluk Poljak and Miroslav Sura, An algorithm for graceful labelling of a class of symmetrical trees, Ars Combin., 14 (1982) 57–66; MR 84d:05150 [1969, 1128].

D. F. Robinson, Comparison of labeled trees with valency three, J. Combin. Theory Ser. B., 11 (1971) 105–119; MR 45 #3243 [1980, 552].

Øystein J. Rødseth, On $h$-bases for $n$ (I &)II, Math. Scand., 48 (1981) 165–183 and 51 (1982) 5–21; MR 82m:10034 and 84c:10019 [1980, 206].

Witold Rzymowski and Adam Stachura, Solution to a problem of guarding territory... (wrc) [1984, 416].

Allen J. Schwenk, How many rinds can a finite sequence of pairs have? Proc. 5th Internat. Conf. Theory Appl. Graphs (Y. Alavi et al., eds.), Wiley, New York, 1985 (tbp) [1982, 113].

Temple F. Smith and Michael S. Waterman, On the similarity of dendrograms, J. Theoret. Biol., 73 (1978) 789–800; MR 58 #15293 [1980, 582].

F. Topsoe, On construction of measures, Math. Inst. Univ. Kobenhavns, Preprint No. 27, 1974 [1984, 562].

Jerrold B. Tunnell, A classical diophantine problem and modular forms of weight 3/2, Invent. Math., 72 (1983) 323–334; MR 85d:11046 [1980, 43].

Peter L. Walker, Iterated complex radicals, Math. Gaz., 67 (1983) 269–273 [1970, 388].

---

# NOTES

EDITED BY SABRA S. ANDERSON, SHELDON AXLER, AND J. ARTHUR SEEBACH, JR.

*For instructions about submitting Notes for publication in this department see the inside front cover.*

## A COMBINATORIAL THEOREM ON CIRCULANT MATRICES

DEAN S. CLARK

*Department of Mathematics, University of Rhode Island, Kingston, RI 02881*

Introductory treatments frequently say that matrices are just rectangular arrays of numbers, but, of course, the real intent is to develop their algebraic properties. Here, we go no further than matrices as formal or symbolic objects. Half the time, the elements will be *colors* rather than numbers. Matrices have a capacity for storing combinatorial information, and we want to exploit that. The following examples extend some classical problems, and provide motivation.

**1. Problem of $n$ queens** (Gauss, ca. 1850). Can $n$ mutually invulnerable queens be placed on an $n \times n$ chessboard? Surprisingly, it seems that the answer—affirmative for $n \geqslant 4$—has been known only since 1969 **[1]**, **[2]**. We give a simple new proof of this fact, below. Here, however, our interest is in other surfaces with chessboard-like properties: the wrapped-around cylindrical or toroidal boards. Whether on the $8 \times 8$, $9 \times 9$, or $10 \times 10$ versions, the enhanced diagonals can make trying to place immune queens a real exercise in futility.

*When can n invulnerable queens be placed on an $n \times n$ toroidal chessboard?*

**2. Ring dancers.** $N$ husbands have joined hands in a large circle, facing inward, while their $N$ wives have linked arms in an inner circle, facing outward. The two groups are dancing in opposite directions. Now, freeze the action. Picking a husband at random, we expect that the person facing him is not his wife. On the other hand, it isn't unreasonable to suspect that, somewhere on the circle, two spouses are facing each other.

*Is it always possible to freeze the action at an instant when no spouses are facing each other?*

**3. $N$ pigeons, $N$ holes.** The pigeonhole principle says that it takes at least $N + 1$ pigeons, flying into $N$ pigeonholes, to guarantee that one of the pigeonholes will contain several pigeons.

Here is a classic application [3], [4], [5], [6]:

Suppose that a disk has been divided into $n = 2k$ congruent sectors with $k$ sectors colored blue and $k$ colored red. Let the same thing be done to a smaller disk. The problem is to show that the smaller disk can always be positioned concentrically within the larger so that at least half the sectors have matching color. We comment on the solution later. For our nonstandard application, which removes the critical $N + 1$st pigeon (see below), let $n$ be arbitrary and, instead of using colors, label the sectors of each disk with the numbers $1, 2, \dots, n$ in any order. We can place the disks concentrically and talk about the sum of the integer labels in a sector.

*Can the disks be positioned so that at least two sector sums are equal?*

**4. A beautiful deception.** To perpetrate this admitted flim-flam, one needs an accomplice seated anonymously in the audience. Participants are asked to draw a clockface with the numbers $1, 2, \dots, 12$ arranged in any order they please. A persuasive MC will stress the secrecy of each individual's arrangement, that there are $11!$ or nearly 40 million possibilities, etc. This being done, everyone is asked to note the largest sum given by three consecutive numbers on their permuted clockface. This becomes a player's *secret number*. Now the bidding starts. A person is to bid any number if they are convinced that no one else in the audience can undercut them with their secret number. For example, I would bid 27 if I felt certain that no one else had a secret number of 26 or less. The winner is either the last bidder or the *first* one who undercuts another bidder. It's a game of nerve and reflexes. Invariably, the accomplice wins.

*What is the accomplice's secret number, and what is the probability that someone else in the audience has it?*

Interestingly, all these examples boil down to questions about the elementary structure of circulant matrices. Recall that a circulant (retrocirculant) is a square matrix identified by its top row. This row is shifted successively a step to the right (left) with wraparound to produce the remaining rows. We use the notation $\mathrm{Circ}(a, b, c, \dots)$ ($\mathrm{Circ}_{\leftarrow}(a, b, c, \dots)$) where $a, b, c, \dots$ are distinct colors. Major (minor) diagonals are the monochrome sets in $\mathrm{Circ}(a, b, \dots)$ ($\mathrm{Circ}_{\leftarrow}(a, b, \dots)$). The toroidal chessboard, ring dancers, and sectored disks are covered by the following basic result.

THEOREM. *It is possible to choose $n$ positions in the $n \times n$ circulant matrix* $\mathrm{Circ}(a, b, c, \dots)$ *which are* (i) *on distinct columns*, (ii) *on distinct rows*, (iii) *of distinct colors*; *if and only if $n \geqslant 1$ is odd. It is possible to replace* (iii) *with* (iii′) *of distinct colors both in* $\mathrm{Circ}(a, b, c, \dots)$ *and* $\mathrm{Circ}_{\leftarrow}(a, b, c, \dots)$; *if and only if $n \geqslant 1$ is an odd integer not divisible by* 3.

*Proof.* Let $n \geqslant 1$ be odd. Superimpose $\mathrm{Circ}(0, 1, 2, \dots, n-1)$ on $\mathrm{Circ}(a, b, \dots)$. Let $(x)_n$ denote the nonnegative remainder after $x$ is divided by $n$. The position in the $k$th column and $a_k$th row shows diagonal number $(k - a_k)_n$, $k = 0, 1, \dots, n-1$. To locate the $n$ positions which satisfy (i)–(iii), begin circling elements starting with the upper leftmost (zero). We move from left to right, as the knight does in chess, over a column and down two rows circling elements as we go. This staircase pattern wraps around at midboard and continues to the last column. The $n$ circled positions are on distinct rows because $n$ is odd. The corresponding diagonal sequence $\{0, n-1, n-2, \dots, 3, 2, 1\}$ shows the circled positions have distinct color in $\mathrm{Circ}(a, b, c, \dots)$.

To argue that $n$ odd is necessary for (i)–(iii), suppose that in column $k$ the $a_k$th row element has been chosen, $k = 0, 1, \dots, n-1$. Since the $a_k$ and $(k - a_k)_n$ are distinct,

$$0 \equiv_n \sum_{k=0}^{n-1} (k - a_k)_n = \sum_{k=0}^{n-1} k = \frac{1}{2}(n-1)n,$$

where $\equiv_n$ denotes congruence modulo $n$. This means that $n$ must be odd.

Now consider the stronger condition (iii′). Superimpose $\mathrm{Circ}_{\leftarrow}(0, 1, 2, \dots, n-1)$ on $\mathrm{Circ}_{\leftarrow}(a, b, \dots)$. Circle the same $n$ positions as before. In general, they will not lie on distinct minor diagonals, as exemplified by the $3 \times 3$ case. But if $n \geqslant 1$ is not divisible by 3, the

corresponding diagonal sequence $\{(k+a_k)_n\}_{k=0}^{n-1}$ will be $\{0,3,6,\dots\}$ without possibility of repetition. These same fixed positions are distinctly colored in Circ$(a,b,\dots)$ and Circ$_{\leftarrow}(a,b,\dots)$ simultaneously.

In fact, $n \not\equiv 0 \pmod 3$ is necessary for (iii′). To see this, suppose that $\{a_k\}$, $\{(k-a_k)_n\}$ and $\{(k+a_k)_n\}$, $0 \leqslant k \leqslant n-1$, are all permutations of $\{0,1,\dots,n-1\}$. Then,

$$\sum_{k=0}^{n-1}(k-a_k)_n^2+\sum_{k=0}^{n-1}(k+a_k)_n^2=2\sum_{k=0}^{n-1}k^2 \text{ implies } 0 \equiv_n \sum_{k=0}^{n-1}a_k^2=\frac{(n-1)(2n-1)n}{6},$$

which holds if and only if $n$ is odd and not divisible by 3.

COROLLARY 1. *It is possible to place $n$ invulnerable queens on the cylindrical or toroidal chessboard if and only if the unwrapped $n\times n$ board satisfies $n \geqslant 5$, $n$ odd and not divisible by* 3.

COROLLARY 2. *It is possible to place $n$ invulnerable queens on an $n\times n$ chessboard for all $n \geqslant 4$.*

Corollary 1 follows by observing that the monochrome diagonals of Circ$(a,b,\dots)$ (Circ$_{\leftarrow}(a,b,\dots)$) model the major (minor) diagonals of a wrapped-around chessboard. Queens placed in the circled positions on such a board remain invulnerable when it is unwrapped. Since a queen has been placed in the upper leftmost square, pruning the first row and column leaves the remaining queens in immune formation. Thus, we have already solved the classical queens problem of Corollary 2 for odd values of $n$, $3 \nmid n$, and the corresponding even values $n-1$.

*Completion of proof of Corollary* 2. Take $n=2m+1>3$, where $3|n$. We modify the construction in [2] slightly and use circulant matrix arguments. As before, the queen placed in the upper leftmost square is deleted along with the first row and column to give the solution for $n-1$. This being done, consider the remaining $2m\times 2m$ submatrix in Circ$_{\leftarrow}(m,m+1,\dots,n-1,0,1,\dots,m-1)$, shown below.

$$\left(\begin{array}{c|ccccc|cccccc}
m & m+1 & m+2 & \cdots & n-2 & n-1 & 0 & 1 & 2 & \cdots & m-2 & m-1 \\
\hline
m+1 & m+2 & m+3 & \cdots & n-1 & 0 & 1 & 2 & 3 & \cdots & m-1 & m \\
 & & & & \ddots & & & & \ddots & & & \\
\vdots & \vdots & & & x & & & & y & & & \vdots \\
 & & & & & \ddots & & & & \ddots & & \\
n-1 & 0 & 1 & \cdots & m-2 & m-1 & m & m+1 & m+2 & \cdots & n-3 & n-2 \\
\hline
\multicolumn{6}{c|}{L} & \multicolumn{6}{c}{R} \\
\hline
0 & 1 & 2 & \cdots & m-1 & m & m+1 & m+2 & m+3 & \cdots & n-2 & n-1 \\
1 & 2 & 3 & & & & & & & & & \\
 & & & \ddots & & & & \ddots & & & & \\
\vdots & \vdots & & w & & & & z & & & & \vdots \\
 & & & & \ddots & & & & \ddots & & & \\
m-1 & m & m+1 & \cdots & n-3 & n-2 & n-1 & 0 & 1 & \cdots & m-3 & m-2
\end{array}\right)$$

Partition this submatrix into left and right halves, denoted $L$ and $R$ respectively. Place the remaining queens in positions circled as follows. Begin by circling 0 in the first column of $L$, then, moving from left to right, one column over and two rows down with $w$ denoting a typical circled position in the lower half of $L$, $w = 3h$, $h = 0, 1, \ldots$. Wrap around to the top of $L$, where $x$ is the typical term, $x = 3i + 1 \pmod{n}$, and continue until all $m$ columns of $L$ are filled. We perform the inverse process in $R$: Circle $n - 1$ in the last column of $R$, then, moving from right to left, over one column and up two rows, with $y$ denoting a typical circled position in the upper half of $R$, $y = n - 3j - 1$, $j = 0, 1, \ldots$. Wrap around to the bottom of $R$, where $z$ is the typical term, $z = n - 3k - 2 \pmod{n}$, and continue until all $m$ columns of $R$ are filled.

Clearly, the queens occupy distinct rows and columns. Because equations like $w = y$ have no solution ($3 \nmid n - 1, 3 \nmid n - 2$), attack along minor diagonals is impossible. On the other hand, $x = z$ does have solutions, and these queens would be vulnerable on the toroidal chessboard. Here on the unwrapped board they are safe in $L$ and $R$.

Confirming immunity along major diagonals is even easier. One checks the same positions in $\mathrm{Circ}(m - 1, m, \ldots, n - 1, 0, 1, \ldots, m - 2)$.

COROLLARY 3. *In Example* (2) *there is always an instant when none of the face-to-face pairs are spouses if and only if $N$ is even. Similarly, in Example* (3), *there is always a superposition of disks with at least two equal sector sums if and only if $n$ is even.*

*Proof.* Example (2) is absorbed into Example (3) by considering the disk problem with $n$ *distinct* colors (= spouses) per disk. If the disks can be superimposed so that no sectors match, the ring dancers problem is solved. Let $s, t, u, \ldots$ denote either the color pattern or sequence of integer labels of the sectors of the larger disk, written linearly. Let $a, b, c, \ldots$ be the same for the smaller disk. Write $\mathrm{Circ}(a, b, c, \ldots)$, the inventory of superpositions, directly underneath $s, t, u, \ldots$. Suppose that $n$ is even. If $s, t, u, \ldots$ are colors, move from left to right circling the element in each column of $\mathrm{Circ}(a, b, \ldots)$ which matches the color of $s, t, u, \ldots$ heading that column. If $s, t, u, \ldots$ are integers, circle $n + 1 - s, n + 1 - t, n + 1 - u, \ldots$.

Think of the circles as pigeons and the rows of $\mathrm{Circ}(a, b, \ldots)$ as the pigeonholes. In the classical (red-blue) disk problem, there would be $k$ pigeons per column (half the column entries are color matches), giving a total of $2k^2$ pigeons filling up $2k = n$ pigeonholes. One of the rows of $\mathrm{Circ}(a, b, \ldots)$ would wind up with at least $k$ circled elements—the desired superposition.

But here, the standard pigeonhole principle doesn't apply. Because the colors are distinct each column gets a single circle, hence, $n$ pigeons into $n$ pigeonholes. Enter the Theorem: The circled positions are in distinct columns and have distinct colors, but $n$ is even. Thus, there must be a row in $\mathrm{Circ}(a, b, \ldots)$ with at least two circled elements. In the sector-sums case, this means that

$$(n + 1 - x) + x = n + 1 = (n + 1 - y) + y,$$

where $x$ and $y$ label two sectors of the larger disk.

For the ring dancers, the pigeonhole principle works in reverse. The two circled elements in one row imply that there must be another row *without* circled elements. That is the instant when no spouses face each other.

All these claims break down when $n$ is odd. In that case, given colors $s, t, u, \ldots$, construct $\mathrm{Circ}(a, b, \ldots)$ by first replicating $s, t, u, \ldots$ from left to right up the minor main diagonal and then extending the colors along major diagonals. Now there is no row of $\mathrm{Circ}(a, b, \ldots)$ without a match with an element of $s, t, u, \ldots$. If $s, t, u, \ldots$ are integers and $n$ is odd, it isn't hard to label the sectors so that every superposition gives distinct sector sums.

Example (4) is connected with circulants in a different way. Let $x_1, x_2, \ldots, x_n$ denote a circular permutation of the integers $1, 2, \ldots, n$, $n \geqslant 7$, and $s_j = x_j + x_{j+1} + x_{j+2}$. Then, $S = \mathrm{Circ}(1, 1, 1, 0, \ldots, 0) X$, where $X^T = (x_1, \ldots, x_n)$, $S^T = (s_1, \ldots, s_n)$. Obviously, not every choice of $S$ can result in a feasible solution $X$ since the elements of $X$ must be distinct integers.

LEMMA. *Nondistinct solutions of* $S = \mathrm{Circ}(1,1,1,0,\ldots,0)X$ *result* (i) *if two consecutive elements in* $S$ *are equal*; *or* (ii) *if* $S$ *contains five consecutive elements of the form abcba.*

*Proof.* The equation $x_j + x_{j+1} + x_{j+2} = x_{j+1} + x_{j+2} + x_{j+3}$ implies $x_j = x_{j+3}$, confirming (i). To confirm (ii), suppose, without loss of generality, that

$$x_1 + x_2 + x_3 = a,\ x_2 + x_3 + x_4 = b,\ x_4 + x_5 + x_6 = b,\ x_5 + x_6 + x_7 = a.$$

These equations imply $x_1 - x_4 = a - b = x_7 - x_4$, so $x_1 = x_7$.

The key question regarding the flim-flam game of Example (4) is: What can a player's secret number possibly be? Obviously it can't be, say, 14 because everyone has a 12 on their clockface and, next to it, at least 1 and 2. The bidding should begin at 16. The bids 17, 18, 19, 20, ... pass uneventfully.

Let $\sigma$ denote any secret number. To obtain a lower bound on $\sigma$ we compute $\bar{s}$, the average value of the triple-sums $s_j$:

$$\bar{s} = \frac{1}{12}\sum_{k=1}^{12} s_k = \frac{1}{4}\sum_{k=1}^{12} x_k = \frac{1}{4}\sum_{k=1}^{12} k = 19.5.$$

Thus, $\sigma \geqslant 20 > \bar{s}$. In actuality, $\sigma$ can never equal 20. Twelve integers, the largest of them 20, average out to 19.5 only if at least 6 of them equal 20. If exactly 6 of the $s_j$ equal 20, then the others must all be 19. In that case, either two consecutive $s_j$ are the same, or else the 20's and 19's alternate. Both possibilities are violations, in view of the above Lemma. If more than 6 of the $s_j$ equal 20, we have the same violation because (the pigeonhole principle!) at least two consecutive entries in $S$ must be 20.

Thus, $\sigma \geqslant 21$. Meanwhile, my accomplice is ready with a circular permutation for which $\sigma$ attains its lower bound. It is 1, 8, 10, 3, 5, 9, 4, 6, 11, 2, 7, 12.

The only remaining question: What is the probability that someone in the audience has also stumbled onto a circular permutation of $1, 2, \ldots, 12$ for which $\sigma = 21$? Frankly, I don't know the probability, but I believe that it is very small. After all, my accomplice and I have never lost a game.

**Acknowledgment.** I wish to thank Professors P. J. Davis and O. Shisha for helpful discussions.

**References**

1. C. Berge, Graphs and Hypergraphs, North-Holland, 1973.
2. E. J. Hoffman, J. C. Loessi, and R. C. Moore, Constructions for the solution of the *m* queens problem, Math. Mag., 42 (1969) 66–72.
3. G. Berman and K. D. Fryer, Introduction to Combinatorics, Academic Press, 1972.
4. D. I. A. Cohen, Combinatorial Theory, Wiley, 1978.
5. C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, 1977.
6. A. Tucker, Applied Combinatorics, Wiley, 1980.

## MULTIPLICATIVE RELATIONS FOR SUMS OF INITIAL $k$TH POWERS

DAVID E. PENNEY AND CARL POMERANCE*
*Department of Mathematics, University of Georgia, Athens, GA 30602*

Is the curious and pretty identity

$$(1 + 2 + \cdots + n)^2 = 1^3 + 2^3 + \cdots + n^3$$

*Supported in part by a National Science Foundation grant.

**C E N T E R   S E C T I O N**
(Vol. 92, No. 10, December 1985)

**Telegraphic Reviews**

Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books submitted for review should be sent to Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.

Telegraphic reviews are designed to give prompt notice of all new books in the mathematical sciences. Certain of these books will be selected for more extensive review in the Reviews section of the Monthly.

Special Codes:

| | | | |
|---|---|---|---|
| T: | Textbook | 13-18: | Grade Level |
| S: | Supplementary Reading | 1-4 : | Time in Semesters |
| P: | Professional Reading | ** : | Special Emphasis |
| L: | Undergraduate Library | ?? : | Questionable |

**General, P*, L.** Perspectives in Mathematics. Ed: W. Jäger, J. Moser, R. Remmert. Birkhauser Boston, 1984, 587 pp, $59.95. [ISBN: 3-7643-1624-1] 22 papers outlining recent developments in diverse areas of mathematics, prepared in honor of the 40th anniversary of the Mathematics Research Institute in Oberwolfach. A splendid sampler of contemporary mathematics. LAS

**Elementary, T(13: 1).** Intermediate Algebra for College Students, Third Edition. Karl J. Smith, Patrick J. Boyle. Brooks/Cole, 1985, xiv + 529 pp. [ISBN: 0-534-03987-1] Usual topics, including polynomials, exponentials and logarithms, conic sections, series and the binomial theorem. Coordinated treatment of equations and inequalities; emphasis on graphing. Attractive format. Problem sets are graded by difficulty. This edition features enlarged collections of review problems. LCL

**Elementary, T(13: 1).** Business Mathematics, Third Edition. Esther H. Highland, Roberta S. Rosenbaum. Reston, 1985, viii + 600 pp. [ISBN: 0-8359-0593-4] Each chapter has been slightly enlarged to include a statement of objectives at the beginning, additional problems (many requiring a calculator) in the middle, and new review problems (with complete solutions) at the end. (First Edition, TR, August-September 1979; Second Edition, TR, March 1982.) LCL

**Mathematics Appreciation, S(13).** Leonardo's Dessert: No Pi. Herbert Wills. NCTM, 1985, iv + 28 pp, $3.50 (P). [ISBN: 0-87353-221-X] For high school teachers and students, an amusing pamphlet dealing with elementary questions of quadrature and equidecomposability of figures (whose boundaries are generally made up of circular arcs) that interested Leonardo da Vinci. SS

**Precalculus, T(13: 1, 2), S.** Functions and Graphs: A Precalculus Course. Raymond A. Barnett. McGraw-Hill, 1985, xv + 765 pp, $31.95. [ISBN: 0-07-003896-1] Presupposes a course in intermediate algebra. Elementary functions and their graphs, trigonometry, conics, systems of equations and inequalities, matrices, and sequences. More concise than authors' other precalculus works. FLW

**Precalculus, T*(13), S.** Before Calculus: Functions, Graphs, and Analytic Geometry. Louis Leithold. Harper & Row, 1985, xi + 730 pp. [ISBN: 0-06-043928-9] A precalculus course which includes polynomial, rational, exponential, logarithmic and trigonometric functions, analytic geometry, sequences, series, systems of equations and complex numbers. Plenty of examples, exercises and illustrations. A well-written text. CEC

**Education, P*, L.** Guidelines for Evaluating Computerized Instructional Materials, Second Edition. William P. Heck, et al. NCTM, 1984, 26 pp, $3 (P). [ISBN: 0-87353-219-8] An updated edition of this valuable set of guidelines. Includes pointers to sources of information, guidelines for evaluating instructional software, guidelines for software documentation, and other useful information. (First Edition, TR, October 1981.) AO

**History, S(17-18), P.** Wolfgang Pauli: Scientific Correspondence with Bohr, Einstein, Heisenberg a.o., V. II: 1930-1939. Ed: Karl von Meyenn, Armin Hermann, Victor F. Weisskopf. Springer-Verlag, 1985, xxxix + 783 pp, $110. [ISBN: 0-387-13609-6] This volume should fascinate theoretical physicists and historians of physics, and interest many others who cannot understand the technical parts. (Volume I, TR, May 1980.) JD-B

**Logic, P.** Consequences of Martin's Axiom. D.H. Fremlin. Tracts in Math., No. 84. Cambridge U Pr, 1984, xii + 325 pp, $54.50. [ISBN: 0-521-25091-9] Martin's axiom is a statement about the cardinality of certain partially ordered sets which is implied by but does not imply the continuum hypothesis. This book presents consequences of Martin's axiom in set theory, combinatorics, topology, measure theory, functional analysis, and group theory. No logic assumed except basic knowledge of cardinals and ordinals. Extensive bibliography. KS

**Foundations, P.** Integration of Science and the Systems Approach. Ed: Zdenek Javurek, Arkadiy Dmitriyevich Ursul, Jirí Zeman. Elsevier Sci, 1984, 257 pp, $55.75. [ISBN: 0-444-99633-8] Marxist-

Leninist philosophy of science, "providing substantiated scientific answers to basic questions of world outlook by generalizing scientific insights from the development of the revolutionary class struggle..." Emphasizes--from this point of view--synthetic trends in scientific cognition and unifying processes of science and philosophy. LAS

**Foundations, T*(16-18: 1, 2), S, L.** Toposes, Triples and Theories. Michael Barr, Charles Wells. Grund. der math. Wissenschaften, B. 278. Springer-Verlag, 1985, xiii + 345 pp, $36. [ISBN: 0-387-96115-1] A well-written introductory text with a reasonable index, a number of exercises with each of nearly 60 sections, and both historical notes and an attempt to provide insight as well as definitions. JAS

**Combinatorics, S(16-18), P.** Geometrical Combinatorics. Ed: F.C. Holroyd, R.J. Wilson. Research Notes in Math., V. 114. Pitman, 1984, 98 pp, $15.95 (P). [ISBN: 0-273-08675-8] What is "geometrical combinatorics?" The answer appears to be: the set of geometric problems that employs combinatorial (and algebraic) techniques for their study together with combinatorial problems that are tightly related to geometry. The current volume consists of eight such papers presented at a 1984 conference on such matters held at the Open University in England. SS

**Combinatorics, P.** Combinatorics on Words: Progress and Perspectives. Ed: Larry J. Cummings. Academic Pr, 1983, x + 405 pp, $34.50. [ISBN: 0-12-198820-1] The proceedings of a meeting held at the University of Waterloo, Canada on August 16-22, 1982. Includes 19 papers. CEC

**Number Theory, T(16: 1), S*, P*, L*.** From Fermat to Minkowski: Lectures on the Theory of Numbers and Its Historical Development. Winfried Scharlau, Hans Opolka. Undergrad. Texts in Math. Springer-Verlag, 1985, xi + 184 pp, $24. [ISBN: 0-387-90942-7] An historically-oriented introduction to number theory including the theory of binary quadratic forms and zeta-functions. The authors begin with Fermat, and trace the development of number theory through the time of Minkowski. Fascinating reading for anyone with an interest in number theory. No exercises. CEC

**Number Theory, S(18), P.** Séminaire de Théorie des Nombres, Paris 1982-83. Ed: Marie-José Bertin, Catherine Goldstein. Prog. in Math., V. 51. Birkhauser Boston, 1984, vi + 312 pp, $29.95. [ISBN: 0-8176-3261-1] Continuing in the tradition of the annual Séminaire de Théorie des Nombres, this volume contains a collection of papers based on lectures presented in 1982-83. A wide variety of topics in number theory is covered in this volume. CEC

**Linear Algebra, T*(13-14: 1).** An Introduction to Linear Algebra with Applications. Steven Roman. Saunders College, 1985, xviii + 500 pp. [ISBN: 0-03-064017-2] For the first course, whether or not it presupposes calculus. Nicely written; attractive format. Very many examples and exercises. Choice of these and of optional sections to cover can steer a course in theoretical or applied direction, as desired. Includes a chapter on computational methods. DFA

**Linear Algebra, T(16-17: 1).** Matrizen und ihre Anwendungen: Teil 1: Grundlagen. Rudolf Zurmühl, Sigurd Falk. Springer-Verlag, 1984, xiv + 342 pp, $33. [ISBN: 0-387-12848-4] First part of a two-volume introduction intended for applied mathematicians, physicists and engineers. Topics and approach fairly standard. No exercises. JD-B

**Algebra, P.** The Curves Seminar at Queen's, Volume III. Ed: Anthony V. Geramita. Papers in Pure & Appl. Math., No. 67. Queen's U, 1984, 269 pp, (P). Contains fourteen seminar talks and a problem section on algebraic curves and commutative rings. LN

**Algebra, P.** Selected Topics in Algebra and its Interrelations with Logic, Number Theory and Algebraic Geometry. Ionel Bucur. Transl: Mihnea Moroianu. Math. & Its Applic. D Reidel, 1984, viii + 406 pp, $79. [ISBN: 90-277-1671-4] This book discusses quadratic forms and the part of algebraic K-theory having to do with "symbols," relations of algebra with logic and algorithmic questions, the theory of topoi and elementary non-standard theory, elliptic functions and curves, part of arithmetic algebraic geometry and rings of continuous functions. For the sophisticated reader. CEC

**Algebra, P.** Galois Groups and Their Representations. Ed: Y. Ihara. Adv. Stud. in Pure Math., No. 2. Elsevier Sci, 1983, ix + 172 pp, $38.50. [ISBN: 0-444-86835-6] Proceedings of a symposium held December 14-18, 1981 in Nagoya, Japan. Focuses on Galois groups of local or global fields, including higher dimensional fields. KS

**Algebra, P.** Lecture Notes in Mathematics-1099: Tame Algebras and Integral Quadratic Forms. Claus Michael Ringel. Springer-Verlag, 1984, xiii + 376 pp, $18 (P). [ISBN: 0-387-13905-2] The aim of these notes is two-fold: to give an introduction to parts of the new representation theory of finite dimensional algebras, and to exhibit the structure of the module categories of a class of algebras having 6, 8, 9, or 10 simple modules. CEC

**Algebra, T(15-16: 2).** Algebra: 2., durchgesehene Auflage. Hans-Jörg Reiffen, Günter Scheja, Udo Vetter. Bibliographisches Inst, 1984, 272 pp, (P). [ISBN: 3-411-05110-8] A brief, clearly-written and largely conventional text for a first course in abstract algebra. Exercises, most of them fairly simple. JD-B

**Algebra, T(17), S, P*.** Quadratic and Hermitian Forms. Winfried Scharlau. Grund. der math. Wissenschaften, No. 270. Springer-Verlag, 1985, x + 421 pp, $48.50. [ISBN: 0-387-13724-6] A sophisticated introduction to some of the main parts of the theory of quadratic and Hermitian forms. The point of view is algebraic, which leads to a number of new results and to some simplifications of classical

results. Includes a good list of references but no exercises. CEC

**Calculus, T**(13: 1).** Calculus I, Second Edition. Jerrold Marsden, Alan Weinstein. Springer-Verlag, 1985, xv + 383 pp, $17.95 (P). [ISBN: 0-387-90974-5] This well-written text contains everything which is desirable in an introduction. Lots of worked examples, excellent graphics, an ample supply of graded exercises, historical remarks, calculator exercises, review sections for precalculus and at the end of chapters, and some interesting applications. This first volume of a three-volume set is meant to be used as a one-semester introduction (First Edition, TR, May 1980). CEC

**Calculus, T*(13-14: 2, 3), S*, L.** Calculus with Analytic Geometry. George F. Simmons. McGraw-Hill, 1985, xxi + 950 pp, $40.95. [ISBN: 0-07-057419-7] A conventional but elegantly and clearly written introductory text. Topics and order of coverage are standard. Development stresses technique over theory; some proofs and deeper theoretical results appear in an appendix. Little treatment of numerical methods. Unusual features: interesting, extended appendices on miscellaneous classical topics, biographical sketches, helpful chapter introductions, liberal arts focus. PZ

**Calculus, T(13-14: 1, 2), S, L.** Mathematical Analysis for Business and Economics, Second Edition. Jagdish C. Arya, Robin W. Lardner. Prentice-Hall, 1985, xv + 764 pp, $31.95. [ISBN: 0-13-561101-6] An introduction in three roughly equal sections to precalculus mathematics, finite mathematics, and elementary calculus. Proof is not stressed. Nearly all applications are to economics. Calculus is developed for rational, logarithmic, and exponential functions only. Little treatment of numerical approximation; few calculator exercises. Includes matrix methods through Markov chains. (First Edition, TR, November 1981.) PZ

**Calculus, T(13: 1, 2).** Calculus with Applications. Coreen L. Mett, James C. Smith. McGraw-Hill, 1985, xiv + 548 pp, $21.95. [ISBN: 0-07-041687-7] For students in business; economics; social, biological, and health sciences. Differentiation and integration and their applications; functions of several variables. Intuitive. Abundantly many examples and exercises from a wide range of areas. Numerical methods, including use of calculator. DFA

**Real Analysis, T(16-17), S, P.** L'Analyse Non Standard. Alain Robert. Papers in Pure & Appl. Math., No. 69. Queen's U, 1984, 119 pp, (P). An introduction to nonstandard analysis for those who wish to progress quickly to the applications, without excessive attention to the logical underpinnings of the subject. The first part covers elementary real analysis (continuity, differentiation, integration), and the second part presents applications to measure theory, approximation theory, differential equations, and Hilbert space. LCL

**Complex Analysis, T(17: 1), S.** Introduction to Complex Analysis. Kunihiko Kodaira. Cambridge U Pr, 1984, ix + 256 pp, $49.50; $17.95 (P). [ISBN: 0-521-24391-2; 0-521-28659-X] An introduction to classical function theory through Riemann mapping and analytic continuation. Unusual feature: the author proves Cauchy's formula first for circular contours only, then derives elementary theory. Topologically general versions of Cauchy's theorems are proved in complete detail later. Many examples and figures; comparatively few exercises. PZ

**Complex Analysis, T(16-17: 1, 2), S, L.** Applied Complex Variables. John W. Dettman. Dover, 1984, ix + 481 pp, $10 (P). [ISBN: 0-486-64670-X] The first half is a fairly quick introduction to the subject, through Riemann mapping and infinite products. Last five chapters, independent of each other, apply theory of the first half to potential theory, ordinary differential equations, Fourier and Laplace transforms, asymptotic expansions. No multivariable calculus is presumed, but some mathematical maturity is. PZ

**Complex Analysis, P.** Lecture Notes in Mathematics-1094: Analyse Complexe. Ed: E. Amar, R. Gay, Thanh Van Nguyen. Springer-Verlag, 1984, ix + 184 pp, $9.50 (P). [ISBN: 0-387-13886-2] Proceedings of the Journées Fermat-Journées SMF, held at Toulouse, May 24-27, 1983. JAS

**Differential Equations, T(16-17), S, P.** Applied Exterior Calculus. Dominic G.B. Edelen. Wiley, 1985, xvii + 471 pp, $44.95. [ISBN: 0-471-80773-7] First half covers essentials of exterior calculus with straightforward computational exercises at end of each chapter. Second half gives applications in second-order partial differential equations, calculus of variations, thermodynamics, electrodynamics, and gauge theory. BH

**Differential Equations, P.** University of Strathclyde Seminars in Applied Mathematical Analysis: Multiparameter Problems. Ed: G.F. Roach. Birkhauser Boston, 1984, viii + 101 pp, £8.50 (P). [ISBN: 1-85014-001-4] A collection of expository papers on linear and nonlinear systems of differential equations involving two or more parameters. AO

**Differential Equations, P.** Differential Equations, Flow Invariance and Applications. N.H. Pavel. Research Note in Math., V. 113. Pitman, 1984, 246 pp, $18.95 (P). [ISBN: 0-273-08651-0] Presentation of recent research on differential equations on closed (or locally closed) sets, notably the theory of flow invariance on Banach manifolds. LAS

**Differential Equations, P.** Differential Equations. Ed: Ian W. Knowles, Roger T. Lewis. Math. Stud., V. 92. Elsevier Sci, 1984, xx + 608 pp, $61.50 (P). [ISBN: 0-444-86875-5] Proceedings of an international conference held at the University of Alabama in Birmingham during March 1983. 76 papers on ordinary and partial differential equations, from participants from 20 countries. LAS

**Differential Equations, P.** Nonlinear Singular Perturbation Phenomena: Theory and Applications. K.W. Chang, F.A. Howes. Appl. Math. Sci., No. 56. Springer-Verlag, 1984, viii + 180 pp, $18.80 (P). [ISBN: 0-387-96066-X] Recent results on the existence and asymptotic behavior of solutions of certain classes of singularly perturbed nonlinear boundary value problems. Both scalar and vector problems are studied using differential inequality techniques. AO

**Partial Differential Equations, T(17-18: 1), S, P.** Partial Differential Equations. Phoolan Prasad, Renuka Ravindran. Wiley, 1984, xi + 252 pp, $24.95. [ISBN: 0-470-20071-5] A basic introduction to partial differential equations. Three chapters treat first-order, linear second-order, and hyperbolic partial differential equations. Physical applications and classical problems are treated as they arise. Good for self-study; each short section contains exercises; most contain worked examples. Limited bibliography. PZ

**Partial Differential Equations, P.** Elliptic Problems in Nonsmooth Domains. P. Grisvard. Mono. & Stud. in Math., V. 24. Pitman, 1985, xiv + 410 pp, $49.95. [ISBN: 0-273-08647-2] Studies elliptic boundary value problems in domains with nonsmooth boundaries and problems with mixed boundary conditions by deriving modified "shift theorems." Presents results mainly in the framework of Sobolev spaces, but also in spaces of Hölder functions. DFA

**Partial Differential Equations, S(16-18), P, L.** Solving Elliptic Problems Using ELLPACK. John R. Rice, Ronald F. Boisvert. Ser. in Comp. Math., No. 2. Springer-Verlag, 1985, x + 497 pp, $46.50. [ISBN: 0-387-90910-9] ELLPACK is a software system for solving elliptic boundary value problems. This volume contains the user's manual, detailed system documentation, and examples illustrating use of the system. AO

**Partial Differential Equations, S(18), P.** Large Deviations and the Malliavin Calculus. Jean-Michel Bismut. Prog. in Math., V. 45. Birkhauser Boston, 1984, viii + 216 pp, $17.95. [ISBN: 0-8176-3220-4] The Malliavin calculus and large deviation techniques are used to study the asymptotics of the conditional probabilities of bridges associated with certain hypoelliptic diffusions. Includes a list of references. CEC

**Partial Differential Equations, P.** Nonlinear Partial Differential Equations and their Applications, Collège de France Seminar, Volume VI. Ed: H. Brezis, J.L. Lions. Res. Notes in Math., No. 109. Pitman, 1984, 323 pp, $27.50 (P). [ISBN: 0-273-08646-4] Written versions of lectures held during the year 1982-1983 at the weekly Seminar in Applied Mathematics at the Collège de France. JAS

**Partial Differential Equations, P.** Lecture Notes in Mathematics-1100: Precise Spectral Asymptotics for Elliptic Operators Acting in Fiberings over Manifolds with Boundary. Victor Ivrii. Springer-Verlag, 1984, v + 238 pp, $11 (P). [ISBN: 0-387-13361-5]

**Numerical Analysis, S(16-17), P.** Least Squares Computations Using Orthogonalization Methods. James W. Longley. Lect. Notes in Pure & Appl. Math., V. 93. Dekker, 1984, xi + 308 pp, $59.75 (P). [ISBN: 0-8247-7232-6] Discusses the use of Gram-Schmidt orthogonalization and Givens transformations for least-squares computations, especially for statistical applications. Listings of FORTRAN implementations of many of the algorithms are included. AO

**Numerical Analysis, P.** Defect Correction Methods: Theory and Applications. Ed: K. Böhmer, H.J. Stetter. Computing Supple., V. 5. Springer-Verlag, 1984, ix + 243 pp, $20 (P). [ISBN: 0-387-81832-4] 13 papers on techniques which use knowledge of some measure of the error to obtain iterative improvements. Includes theoretical foundations; applications to eigenvalue problems, stiff ordinary differential equations, and partial differential equations; multi-grid methods and inclusion algorithms; and adaptive software. RWN

**Numerical Analysis, S(16-17), L*.** Methods of Numerical Integration, Second Edition. Philip F. Davis, Philip Rabinowitz. Comp. Sci. & Appl. Math. Academic Pr, 1984, xiv + 612 pp, $52. [ISBN: 0-12-206360-0] An updated and expanded edition of an important reference. Major changes have been made in the chapters discussing the evaluation of multi-dimensional integrals and automatic integration. (First Edition, TR, January 1976.) AO

**Numerical Analysis, T(16-17: 1), L.** Finite Element Solution of Boundary Value Problems: Theory and Computation. O. Axelsson, V.A. Barker. Comp. Sci. & Appl. Math. Academic Pr, 1984, xviii + 432 pp, $59. [ISBN: 0-12-068780-1] An introduction to both the theoretical and computational aspects of the finite element method for solving boundary value problems for partial differential equations. Primary attention is given to second-order, linear, self-adjoint problems. AO

**Numerical Analysis, T(13-14: 1), L*.** Computing in Applied Science. William J. Thompson. Wiley, 1984, xxiv + 325 pp, $26.95. [ISBN: 0-471-09355-6] An introduction to a number of mathematical topics important in applications (complex numbers, finite difference methods, power series, Fourier expansions, differential equations, vector dynamics, and data analysis) together with related numerical methods. AO

**Functional Analysis, S(18), P.** The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. C.W. Groetsch. Research Notes in Math., V. 105. Pitman, 1984, 104 pp, $16.95 (P). [ISBN: 0-273-08642-1] Motivates material well with short introductory chapter on equations of the first kind, ill-posed problems, generalized inverses, and linear operators in Hilbert space. Then discusses general regularization, Tikhonov regularization and finite approximation methods for Fredholm equations of the first kind. BH

**Functional Analysis, T*(18).** A Course in Functional Analysis. John B. Conway. Grad. Texts in Math., No. 96. Springer-Verlag, 1985, xiv + 404 pp, $38. [ISBN: 0-387-96042-2] Well-motivated graduate text. Assumes measure and integration theory and point set topology and requires at least a concurrent course in analytic function theory. Many good problems of varying difficulty at the end of each section. BH

**Functional Analysis, S(18), P.** Analytic Sets in Locally Convex Spaces. Pierre Mazet. Math. Stud., V. 89. Elsevier Sci, 1984, ix + 275 pp, $38.50 (P). [ISBN: 0-444-86867-4] This book gives definitions and basic tools for the geometry of analytic sets in the infinite dimensional case, and generalizes the theorems on local representation of analytic spaces together with the great classical theorems viz. the Nullstellensatz, the direct image theorem and the theorem of Remmert-Stein. MU

**Functional Analysis, P.** Fifteen Papers on Functional Analysis. Ed: Lev. J. Leifman. AMS Transl. Ser. 2, V. 124. AMS, 1984, viii + 183 pp, $60. [ISBN: 0-8218-3085-6]

**Analysis, T(18: 1), P.** Brownian Motion and Martingales in Analysis. Richard Durrett. Wadsworth, 1984, xi + 328 pp, $34.95. [ISBN: 0-534-03065-3] Brownian motion can be used to prove many results in classical analysis, but most such results are not widely known. The purpose of this book is to bring such results together. Chapters 1 and 2 introduce Brownian motion and the stochastic integral, and the final seven chapters give applications. MT

**Analysis, T(17), S, P.** Lecture Notes in Mathematics-1078: Integration Theory. A.J.E.M. Janssen, P. van der Steen. Springer-Verlag, 1984, 224 pp, $11 (P). [ISBN: 0-387-13386-0] Study of Lebesgue integration which compares the methods of Bourbaki, Daniell, and Carathéodory. Demonstrates that these methods have a large common ground which can be subsumed in a single theory. Clear, well-motivated exposition. Many problems. BH

**Analysis, T(16-18: 1, 2), S, P, L.** Singularities of Differentiable Maps, V. I: The Classification of Critical Points, Caustics and Wave Fronts. V.I. Arnold, S.M. Gusein-Zade, A.N. Varchenko. Mono. in Math., V. 82. Birkhauser Boston, 1985, x + 382 pp, $44.95. [ISBN: 0-8176-3187-9] An example-driven "zoology" of singularities of differentiable maps, covering basic concepts and applications with minimum theoretical or technical details. Adapted from courses given at Moscow State University from 1966 to 1978; translated from the Russian original. LAS

**Analysis, P*.** n-Widths in Approximation Theory. Allan Pinkus. Erge. der Math. und ihrer Grenzgebiete, Band 7. Springer-Verlag, 1985, x + 291 pp, $39. [ISBN: 0-387-13638-X] Develops basic theory of n-widths and applies it to Hilbert spaces, integral operators, matrices, Sobolev spaces, and spaces of analytic functions. Well motivated. Extensive bibliography. BH

**Analysis, P.** Lecture Notes in Mathematics-1108: Global Analysis--Studies and Applications I. Ed: Yu. G. Borisovich, Yu. E. Gliklikh. Springer-Verlag, 1984, 301 pp, $16 (P). [ISBN: 0-387-13910-9] Translations of main articles from the series "New Developments in Global Analysis" (in Russian) published 1982-1984. JAS

**Analysis, P.** Approximation Theory and Spline Functions. Ed: S.P. Singh, J.W. H. Burry, B. Watson. NATO ASI Ser. C., V. 136. D Reidel, 1984, ix + 485 pp, $69.50. [ISBN: 90-277-1818-0] The proceedings of a NATO Advanced Study Institute held at Memorial University of Newfoundland in 1983. Papers cover a wide range of topics: multivariate approximation, spline functions, rational functions, applications of elliptic integrals and functions, and Padé approximation. AO

**Analysis, T*(16-18: 1, 2), S, L.** Curves and Singularities: A Geometrical Introduction to Singularity Theory. J.W. Bruce, P.J. Giblin. Cambridge U Pr, 1984, xii + 222 pp, $39.50; $15.95 (P). [ISBN: 0-521-24945-7; 0-521-27091-X] An elementary (undergraduate) treatment of geometrical applications of singularity theory that grow out of transversality and unfolding. Many examples and exercises make this ideally suited to class or seminar use. LAS

**Analysis, P.** Lecture Notes in Mathematics-1096: Théorie du Potentiel. Ed: G. Mokobodzki, D. Pinchon. Springer-Verlag, 1984, ix + 582 pp, $27.50 (P). [ISBN: 0-387-13894-3] Proceedings of the Colloque Jacques Deny held at Orsay, June 20-23, 1983. JAS

**Analysis, P.** Minimal Surfaces and Functions of Bounded Variation. Enrico Giusti. Mono. in Math., V. 80. Birkhauser Boston, 1984, xii + 240 pp, $39.95. [ISBN: 3-7643-3153-4] "Account of the theory of parametric and non-parametric minimal surfaces of codimension one in Euclidean spaces of arbitrary dimension." Topics include: existence and regularity almost everywhere of solutions to the Plateau problem, minimal cones, dimension of the singular set of surfaces, Dirichlet problem for the minimal surface equation, Bernstein problem for entire solutions. BH

**Algebraic Geometry, P.** Surfaces with Canonical Hyperplane Sections. D.H.J. Epema. CWI Tract, V. 1. Math Centrum, 1984, ix + 105 pp, Dfl. 15,50 (P). [ISBN: 90-6196-267-6] Let X be a projective algebraic surface. X has canonical hyperplane sections if X can be embedded in some projective space in such a way that a general hyperplane section is a smooth canonically embedded curve. Author classifies such surfaces and studies specifically those which are birational to an irrational ruled surface. BH

**Algebraic Geometry, P.** Lecture Notes in Mathematics-1101: Resolution of Surface Singularities. Vincent Cossart, Jean Giraud, Ulrich Orbanz. Springer-Verlag, 1984, vii + 132 pp, $7.50 (P). [ISBN: 0-387-13904-4] The three authors give lectures on the work of Zariski, Hironaka and Abhyankar on the

resolution of the singularities of algebraic surfaces. Emphasis is on easy canonical processes. There is also an appendix by Hironaka on the resolution of excellent surfaces. LN

**Algebraic Geometry, P.** Lecture Notes in Mathematics-1092: Complete Intersections. Ed: S. Greco, R. Strano. Springer-Verlag, 1984, vii + 299 pp, $16 (P). [ISBN: 0-387-13884-6] Lectures given at the First 1983 Session of the Centro Internationale Matematico Estivo (CIME) held at Acireale (Catania), Italy, June 13-21, 1983. JAS

**Geometry, S(15-17), P.** Elementargeometrie: Das Fundament der Analytischen Geometrie. Paul Lorenzen. Bibliographisches Inst, 1984, 238 pp, (P). [ISBN: 3-411-00400-2]

**Differential Topology, P.** Lecture Notes in Mathematics-1102: Stratified Mappings--Structure and Triangulability. Andrei Verona. Springer-Verlag, 1984, ix + 160 pp, $9.50 (P). [ISBN: 0-387-13898-6] A proof that any proper, topologically stable smooth map from one smooth manifold without boundary to another is triangulable. JAS

**Topology, T(15-16: 1, 2), L.** General Topology. Jacques Dixmier. Undergrad. Texts in Math. Springer-Verlag, 1984, x + 140 pp, $18. [ISBN: 0-387-90972-9] Originally written for students in the third university year in France, the present book is a translation of the original together with a few minor improvements and the addition of a section on normal spaces. It is informal but dense by American standards; many proofs are left to the student. Filter base notions are used for definitions of limits and continuity. The exercises for all chapters are ensconced in an unconscionably short exercise section at the end of the book. Includes a chapter on normed spaces culminating in Riesz´s theorem, and a chapter on infinite sums culminating with certain summable families in Hilbert spaces. PH

**Operations Research, T(15), S, L.** Simulation of Waiting-Line Systems. Susan L. Solomon. Prentice-Hall, 1983, xi + 452 pp, $32. [ISBN: 0-13-810044-6] Intended as an introductory text for simulation. Applications deal with waiting-line systems, but techniques apply to other simulations as well. The text includes three chapters discussing use of the simulation package GPSS. MT

**Optimization, S(18), P.** Variational and Quasivariational Inequalities: Applications to Free Boundary Problems. Claudio Baiocchi, António Capelo. Transl: Lakshmi Jayakar. Wiley, 1984, ix + 452 pp, $57.95. [ISBN: 0-471-90201-2] Studies variational and quasivariational inequalities of elliptic type, including the mathematical theory behind them and their application to free-boundary problems. Accessible to anyone with basic understanding of Lebesgue integration theory, measure theory and elementary properties of Hilbert and Banach spaces. No problems. Useful as a reference. Extensive bibliography. BH

**Optimization, T(15-16), S, L.** Introduction to Non-Linear Optimization. L.E. Scales. Springer-Verlag, 1985, xi + 243 pp, $19.80 (P). [ISBN: 0-387-91252-5] A survey of non-linear optimization methods emphasizing those that are generally applicable, widely available, and supported by theory. Unfortunately, no exercises accompany the text. AO

**Optimization, T*(14-16: 1), L*.** Combinatorial Optimization for Undergraduates. L.R. Foulds. Undergrad. Texts in Math. Springer-Verlag, 1984, xii + 227 pp, $36. [ISBN: 0-387-90977-X] An introduction to combinatorial optimization techniques and applications. Topics include linear programming, integer programming, dynamic programming, and network algorithms. AO

**Optimization, P.** Mathematical Programming. Ed: Richard W. Cottle, Milton L. Kelmanson, Bernard Korte. Elsevier Sci, 1984, viii + 357 pp, $59.50. [ISBN: 0-444-86821-6] Proceedings of the International Conference on Mathematical Programming held in Rio de Janeiro in April 1981. Contains twenty-one papers. AO

**Optimization, T(17-18: 1), P.** Network Flows and Monotropic Optimization. R.T. Rockafellar. Wiley, 1984, xiii + 616 pp, $44.95. [ISBN: 0-471-88078-7] Presents an introduction to the theory of network programming problems and generalizations of these results to a larger class of monotropic programming problems (problems in which a preseparable convex function is minimized subject to linear constraints). AO

**Optimization, P.** Variational Convergence for Functions and Operators. H. Attouch. Appl. Math. Ser. Pitman, 1984, 423 pp, $29.95 (P). [ISBN: 0-273-08583-2] "Focuses on minimization problems and develops a convergence theory for sequences of functions called epi-convergence, which may be regarded as the ´weakest´ notion which allows approach to the limit in these problems." Extensive bibliography. BH

**Optimization, T(18), S, P.** Nonlinear Functional Analysis and its Applications III: Variational Methods and Optimization. Eberhard Zeidler. Transl: Leo F. Boron. Springer-Verlag, 1985, xxii + 662 pp, $59. [ISBN: 0-387-90915-X] Translation of the revised and extended German version. Topics include existence and uniqueness principles, extremal problems with and without side conditions, saddle points and duality, and variational inequalities. Useful reference with very extensive bibliography. Numerous problems of varying degrees of difficulty with solutions or references to solutions. BH

**Dynamical Systems, P.** Dynamical Systems and Microphysics: Control Theory and Mechanics. Ed: Austin Blaquière, George Leitmann. Academic Pr, 1984, x + 403 pp, $29.50. [ISBN: 0-12-104365-7] A collection of papers that explore some topics of common interest in contemporary mechanics and mathemati-

cal system theory. AO

**Dynamical Systems, P.** Fluids and Plasmas: Geometry and Dynamics. Ed: Jerrold E. Marsden. Contemp. Math., V. 28. AMS, 1984, xvi + 448 pp, $35 (P). [ISBN: 0-8218-5028-8] Proceedings of the summer research conference held at the University of Colorado, Boulder, in July 1983. The papers discuss geometric and dynamical systems aspects of fluid and plasma dynamics as well as analytical and numerical methods. AO

**Control Theory, P.** Decision and Control in Uncertain Resource Systems. Marc Mangel. Math. in Sci. & Eng., V. 172. Academic Pr, 1985, xiii + 255 pp, $39.50. [ISBN: 0-12-468720-2] A graduate-level textbook, as well as a resource book, for deciding optimal policies on what to do, and how to do, exploration and exploitation of natural resources (exhaustible and renewable). The first two chapters on mathematics aim to keep the book as self-contained as possible, and includes a discussion of optimal stochastic control, one of the major tools used in the specific resource applications which follow. A final chapter is on numerical methods. LCL

**Control Theory, P.** Lecture Notes in Mathematics-1107: Nonlinear Analysis and Optimization. Ed: C. Vinti. Springer-Verlag, 1984, v + 214 pp, $11 (P). [ISBN: 0-387-13903-6] Proceedings of the international conference in honor of Lamberto Cesari held in Bologna, Italy, May 3-7, 1982. JAS

**Systems Theory, P.** System Theoretic Description of Physical Systems. A.J. van der Schaft. CWI Tract, No. 3. Math Centrum, 1984, xv + 256 pp, Dfl. 36,90 (P). [ISBN: 90-6196-269-2] Introduces and describes general system-theoretic models, including finite dimensional linear and nonlinear systems, and applies them to Hamiltonian systems, symmetries and conservation laws, gradient systems and optimal control. Well motivated. BH

**Probability, P.** Advances in Probability Theory: Limit Theorems & Related Problems. Ed: A.A. Borovkov. Optimization Software (Worldwide Distr: Springer-Verlag), 1984, xiv + 377 pp, $48. [ISBN: 0-387-90945-1] Collection of articles on limit theorems grouped into three parts: 1) Invariance principle of Donsker and Prokhorov, 2) Limit theorems for random processes and their applications, 3) Miscellaneous. English translation from Russian. BH

**Probability, T(18: 1), S.** Introduction to Probability Theory. Kiyosi Itô. Cambridge U Pr, 1984, x + 213 pp, $32.50. [ISBN: 0-521-26418-9] A measure-theoretic probability theory text. Topics include probability spaces, random variables and vectors, independence, sigma-algebras, characteristic functions, conditional probability measures, and limit theorems. MT

**Probability, P.** Comparison Methods for Queues and Other Stochastic Models. Dietrich Stoyan. Wiley, 1983, xiii + 217 pp, $39.95. [ISBN: 0-471-10122-2] Presents unified methods for approximating stochastic models by other, simpler models, and for obtaining bounds on stochastic models. The bounds and approximation methods depend on monotonicity and continuity properties of stochastic models. MT

**Probability, T*(16: 1, 2), S, P, L.** Introduction to Probability Models, Third Edition. Sheldon M. Ross. Academic Pr, 1985, xiv + 502 pp, $28. [ISBN: 0-12-598463-4] Revision of the author's 1980 Second Edition (TR, January 1982). The chapter on statistical estimation has been removed and a chapter on Brownian motion and stationary processes and an extensive chapter on simulation have been added. In addition, new material has been added to all but the first chapter. RSK

**Probability, T(15: 1), L.** Introductory Probability Theory. Janos Galambos. Probability: Pure & Appl. Dekker, 1984, vii + 200 pp, $25. [ISBN: 0-8247-7179-6] A text for a one-semester introduction to probability theory. Assumes a basic knowledge of calculus. Covers the usual topics, including conditional probability, independence, random variables, expectation, and the Central Limit Theorem. Separates more involved mathematics from concepts and keeps the mathematical level relatively low. MT

**Probability, S(17), P.** General Irreducible Markov Chains and Non-negative Operators. Esa Nummelin. Tracts in Math., V. 83. Cambridge U Pr, 1984, xi + 156 pp, $37.50. [ISBN: 0-521-25005-6] Deals with theory of general irreducible Markov chains and its connection with the Perron-Frobenius theory of non-negative operators. Applications include queueing theory, storage theory, autoregressive processes and renewal theory. Extensive bibliography. BH

**Probability, P.** Lecture Notes in Mathematics-1097: École d'Été de Probabilités de Saint-Flour XII--1982. R.M. Dudley, H. Kunita, F. Ledrappier. Springer-Verlag, 1984, x + 396 pp, $20 (P). [ISBN: 0-387-13897-8] Three sets of lectures: "A Course on Empirical Processes," by R.M. Dudley; "Stochastic Differential Equations and Stochastic Flow of Diffeomorphisms," by H. Kunita; and "Quelques Proprietes des Exposants Caracteristiques," by F. Ledrappier. JAS

**Probability, T(18: 1).** Martingales and Stochastic Integrals. P.E. Kopp. Cambridge U Pr, 1984, xi + 202 pp, $29.95. [ISBN: 0-521-24758-6] Theoretical presentation of three main topics, discrete time martingales, continuous-time martingales, and stochastic integrals, preceded by introductory material on measure theoretic probability and functional analysis. Also includes an appendix on a non-commutative extension of stochastic integration. RSK

**Probability, T(18: 1), P.** Convergence of Stochastic Processes. David Pollard. Ser. in Stat. Springer-Verlag, 1984, xiv + 215 pp, $28. [ISBN: 0-387-90990-7] This book deals with empirical processes and their convergence properties, discussing convergence under different metrics and on several spaces. MT

**Statistics, S(17), P.** An Introduction to Latent Variable Models. B.S. Everitt. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1984, viii + 107 pp, $20. [ISBN: 0-412-25310-0] This book introduces models involving latent variables, concepts that cannot be observed directly but which have effects on measurable variables. Latent variables are useful mostly in studies in social sciences. MT

**Statistics, P*.** Time Series in the Frequency Domain. Ed: D.R. Brillinger, P.R. Krishnaiah. Handbook of Stat., V. 3. Elsevier Sci, 1983, xiv + 485 pp, $88.50. [ISBN: 0-444-86726-0] Third volume in this series of reference books on statistical methodology and applications (Volume 1, Analysis of Variance, TR, June-July 1981; Volume 2, Classification, Pattern Recognition and Reduction of Dimensionality, TR, December 1983). Contains 20 chapters providing an up-to-date picture of the spectrum approach to time series analysis. RSK

**Statistics, T(14-16: 1, 2), S.** Probability and Statistics for Engineers and Scientists, Third Edition. Ronald E. Walpole, Raymond H. Myers. Macmillan, 1985, xiii + 639 pp. [ISBN: 0-02-424170-9] This new edition contains new examples using real data, several rewritten sections, a discussion of stem and leaf diagrams, new material on residues in regression, Latin square designs, Bartlett's test for equality of several variances, and p values. Presupposes calculus. (First Edition, TR, August-September 1972; Second Edition, TR, December 1978.) FLW

**Statistics, T(16: 1), S.** Elements of Simulation. Byron J.T. Morgan. Chapman & Hall, 1984, xiii + 351 pp, $18.95 (P). [ISBN: 0-412-24590-6] Beginning with a review of basic probability theory results, this book continues with discussion of generating specific random variables, including uniform, normal, exponential, binomial and Poisson random variables. It also includes testing of the randomness of random numbers, variance reduction, and model construction. MT

**Statistics, T(16: 1), S.** Advanced Methods of Data Exploration and Modelling. B.S. Everitt, G. Dunn. Heinemann Educ Books, 1983, xix + 253 pp, $20. [ISBN: 0-435-82294-2] Intended as an intermediate text for those interested in analysis of social or behavioral data. Assumes a first course in statistics and some familiarity with matrix algebra. Includes plotting techniques, multivariate techniques, discussion of the general linear model and latent variable models. MT

**Statistics, T??(13: 1), S.** Statistical Programs in BASIC. Ronald D. Schwartz, David T. Basso. Reston, 1985, xiv + 208 pp, (P). [ISBN: 0-8359-7106-6] Claimed to be a text, this book is a collection of BASIC programs that do calculations for topics which would appear in an introductory statistics course. The format of the book (for each technique) is rigid: statement of problem, algorithm, BASIC program, example problems. MT

**Statistics, S(18), P.** The Forecasting Accuracy of Major Time Series Methods. S. Makridakis. Wiley, 1984, viii + 301 pp, $39.95. [ISBN: 0-471-90327-2] Assesses accuracy of forecasting methods based on an empirical study. The study was a forecasting competition in which participants were asked to forecast up to eighteen time horizons ahead. The book concludes that there are no methods that are clearly best and that accuracy depends on the type of data and the forecasting situation considered. MT

**Statistics, P.** Statistical Methods for Cancer Studies. Ed: Richard G. Cornell. Statistics, V. 51. Dekker, 1984, x + 479 pp, $59.50. [ISBN: 0-8247-7169-9] The specialized nature of studies of cancer has led to development of unique statistical methods. This is a collection of articles describing some of these methods. MT

**Computer Literacy, S(13-16), P.** Microcomputers and their Commercial Applications. D.E. Avison. Comp. Sci. Texts. Computer Science Pr, 1983, viii + 94 pp, $9.95 (P). [ISBN: 0-632-01172-6] A substantial, though non-theoretical, overview of the issues involved in "What's it good for?" and "How do I decide?" Although it is a bit specialized for a computer literacy book, it has much more substance than most of the computer literacy genre. However, it is written as a reference volume rather than as a text. JAS

**Computer Literacy.** First Book on UNIX for Executives. Yukari Shirota, Tosiyasu L. Kunii. Springer-Verlag, 1984, xi + 154 pp, $16 (P). [ISBN: 0-387-70003-X] A strange mixture of hype and cartoons, sure to insult the intended audience. This overview of UNIX may whet a junior high reader's appetite, but the bill of fare is junk food of no lasting value. LAS

**Computer Programming, T(14-16), S, L.** Lisp Programming. I. Danicic. Comp. Sci. Texts. Computer Science Pr, 1983, ix + 96 pp, $11.95 (P). [ISBN: 0-632-01181-5] A serious study of the syntax of LISP focusing on LISP 1.5. This is not a tutorial, but rather a presentation and explanation of the structure and functions of LISP. JAS

**Computer Programming, S(14-15), L.** C: An Advanced Introduction. Narain Gehani. Computer Science Pr, 1985, xv + 332 pp, $29.95. [ISBN: 0-88175-053-0] An excellent introduction to the C programming language for experienced programmers. Compared to more traditional introductory texts, this book places greater emphasis on advanced topics such as data abstraction, exception handling, and concurrent programming. AO

**Computer Programming, T(13: 1).** Prelude to Programming: Problem Solving and Algorithms. William Mitchell. Reston, 1984, ix + 173 pp, $22.95. [ISBN: 0-8359-5614-8] Problem solving strategies and programming methodologies for would-be programmers. An introductory survey of useful techniques: graphical representations of algorithms (flowcharts and Nassi-Shneiderman diagrams), algorithm verification techniques, and efficiency measures. AO

**Computer Programming, S(13-14).** Learning to Program in C. Thomas Plum. Prentice-Hall, 1983, vi + 349 pp, $32.95. [ISBN: 0-13-527854-6] An introductory textbook on the C programming language for novice programmers. Emphasizes techniques that lead to readable, portable code. Not a complete guide to the language. AO

**Computer Programming, S(13-14).** The C Puzzle Book. Alan R. Feuer. Prentice-Hall, 1982, viii + 171 pp, $18.95. [ISBN: 0-13-109934-5] A collection of exercises designed to test the reader's mastery of C language semantics. Useful as a supplement to any textbook on the language. AO

**Computer Programming, T(14-15), S.** IBM-PC 8088 MACRO Assembler Programming. Dan Rollins. Macmillan, 1985, xxiii + 435 pp, (P). [ISBN: 0-02-403210-7] The 8088 processor is the microprocessor inside the IBM-PC. For this reason, it is one of the most popular microprocessors in use today. The text explains the design and architecture of the 8088, as well as how to program in 8088 assembly code. Before reading this, an individual should already know how to program in one or more high-level languages. MS

**Data Structures, T(16-17: 1-3), L.** Data Structures and Algorithms. Kurt Mehlhorn. EATCS Mono. on Theoretical Comp. Sci. Springer-Verlag, 1984, $17.50 each. 1: Sorting and Searching, xiv + 336 pp [ISBN: 0-387-13302-X]; 2: Graph Algorithms and NP-Completeness, xii + 260 pp [ISBN: 0-387-13641-X]; 3: Multi-dimensional Searching and Computational Geometry, xii + 284 pp. [ISBN: 0-387-13642-8] State-of-the-art algorithms are presented and analyzed. Volumes 2 and 3 assume familiarity with some material from Volume 1, but are independent of each other. AO

**Software Systems, S(15-18), P.** TK!Solver for Engineers. Victor E. Wright. Reston, 1984, xii + 419 pp, (P). [ISBN: 0-8359-7711-0] TK!Solver is a program designed to let a user communicate with a computer, without programming, in terms familiar to an engineer. This book is a user's guide for engineers who want to use a computer as a tool, who may not know much about computers, and who would appreciate a thorough guide to the non-trivial tool (program) TK!Solver. JAS

**Software Systems, S**(13-16), L.** Unix for Users. Chris Miller, Roger Boyle. Comp. Sci. Texts. Computer Science Pr, 1984, x + 210 pp, $12.95 (P). [ISBN: 0-632-01182-3] An introduction to UNIX Version 7, intended to be sufficiently generic to be of use with most common versions of UNIX and even UNIX look-alikes. Well written; starts at the beginning and covers a lot of ground. The emphasis is on explaining the nature of UNIX and its standard utilities; details are to be found in the UNIX Programmer's Manual. JAS

**Computer Science, S(16), P.** Frontiers in Computer Graphics. Ed: Tosiyasu L. Kunii. Springer-Verlag, 1985, xi + 443 pp, $45. [ISBN: 0-387-70004-8] The 30 papers in this text were presented at the Computer Graphics Conference in Tokyo, Japan, April 1984. The text is divided into 8 sections with 2-5 papers in each section: geometric modelling, graphics languages, visualization techniques, human factors issue, interactive design, CAD/CAM, display hardware, and standardization issues. In addition to the papers, there are over 250 full-color photographs. MS

**Computer Science, P.** Algorithm Design for Computer System Design. Ed: G. Ausiello, M. Lucertini, P. Serafini. Springer-Verlag, 1984, 236 pp, $18.30 (P). [ISBN: 0-387-81816-2] A collection of nine papers from the July 1983 summer school sponsored by the International Center for Mechanical Sciences. These papers discuss "Combinatorial Problems in Computer System Design" and the "Optimal Design of Parallel Computing Systems." AO

**Computer Science, T(16-18: 1), P, L*.** Introduction to Logic Programming. Christopher John Hogger. APIC Stud. in Data Proc., No. 21. Academic Pr, 1984, xiii + 278 pp, $46. [ISBN: 0-12-352090-8] A comprehensive introduction to the use of symbolic logic as a programming language. In addition to providing an introduction to logic programming, it also discusses the derivation of logic programs from logic specifications and implementation of the logic programming language PROLOG. AO

**Computer Science, S(14), L*.** Modula-2 for Pascal Programmers. Richard Gleaves. Springer-Verlag, 1984, x + 145 pp, $16.95 (P). [ISBN: 0-387-96051-1] A brief introduction to the programming language Modula-2 for persons already familiar with Pascal. Although features unique to Modula-2 are included, the most useful part of the book is the detailed description of the differences from Pascal. The book does not provide a complete introduction to Modula-2. AO

**Computer Science, S(13), L**.** Pascal User Manual and Report: Revised for the ISO Pascal Standard, Third Edition. Kathleen Jensen, Niklaus Wirth. Springer-Verlag, 1985, xvi + 266 pp, $14 (P). [ISBN: 0-387-96048-1] The original text has been modified to reflect the ISO Pascal standard. In addition, Wirth's EBNF syntactic notation is now used and several of the example programs in the user's manual have been improved. An appendix summarizes the changes made in the report (Second Edition, TR, December 1976). AO

**Computer Science, P.** Lecture Notes in Computer Science-182: STACS 85. Ed: K. Mehlhorn. Springer-Verlag, 1985, vii + 374 pp, $16 (P). [ISBN: 0-387-13912-5] The 37 papers in this text were presented at the Second Annual Symposium on Theoretical Computer Science held in Saarbrücken, Germany, January 1985. These papers address a wide range of research issues in theoretical computer science, including computability, formal language theory, coding theory, tree automata, and formal semantics. The papers are all at a very advanced level of presentation. MS

**Computer Science, T(15-17: 1).** A First Course in Formal Language Theory. V.J. Rayward-Smith. Comp. Sci. Texts. Computer Science Pr, 1983, xiii + 123 pp, $14.95 (P). [ISBN: 0-632-01176-9] A short and

to-the-point presentation of regular languages, context-free languages, automata, and parsing. Index (short), chapter exercises, no bibliography. Enjoyable to read and mostly language independent--uses, late in the book, a Pascal-like representation of a few algorithms and only a few problems ask for a program in Pascal. JAS

**Computer Science, S(16-17), P, L*.** The ADA Programming Language: A Tutorial. Sabina H. Saib, Robert E. Fritz. IEEE Computer Society, 1983, ix + 538 pp, $36 (P). [ISBN: 0-8186-0456-5] An anthology containing most of the major early papers on the Ada programming language together with several tutorial papers that describe particular aspects of the language. AO

**Computer Science, T(16-18: 1), L.** The Theory of Relational Databases. David Maier. Computer Science Pr, 1983, xv + 637 pp, $28.95. [ISBN: 0-914894-42-0] A textbook for a second course on databases. Designed to bring the reader almost to the frontier of current research on the subject. AO

**Applications, S(17-18), P.** Mathematische Grundlagen der Kartographie: 2., überabeitete und erweiterte Auflage. Josef Hoschek. Bibliographisches Inst, 1984, 210 pp, (P). [ISBN: 3-411-01661-2] Treats the most important types of map projection, and the mathematics behind them. JD-B

**Applications (Artificial Intelligence), T(16-17: 1), S, P, L.** Building Expert Systems. Ed: Frederick Hayes-Roth, Donald A. Waterman, Douglas B. Lenat. Addison-Wesley, 1983, xvi + 444 pp, $34.95. [ISBN: 0-201-10686-8] A collection of ten articles covering the design, construction, and evaluation of expert systems. A good, not too technical, introduction to the subject. AO

**Applications (Artificial Intelligence), P.** Robotics and Artificial Intelligence. Ed: Michael Brady, Lester A. Gerhardt, Harold F. Davidson. NATO ASI Ser. F, V. 11. Springer-Verlag, 1984, xvii + 693 pp, $62.50. [ISBN: 0-387-12888-3] A collection of invited papers mostly from the NATO Advanced Study Institute held in June 1983 at Castelvecchio Pascoli, Italy. JAS

**Applications (Economics), P.** Lectures on Schumpeterian Economics. Ed: Christian Seidl. Springer-Verlag, 1984, ix + 219 pp, $21.50. [ISBN: 0-387-13290-2] Thirteen lectures on Schumpeter and his economic theories; nine given during Schumpeter centenary celebration at the University of Graz, Austria in 1983. Contributors include six Austrian economists and Peter Hammond of Stanford. KS

**Applications (Economics), P.** Lecture Notes in Economics and Mathematical Systems-232: Bayesian Full Information Analysis of Simultaneous Equation Models Using Integration by Monte Carlo. Luc Bauwens. Springer-Verlag, 1984, vi + 114 pp, $8.50 (P). [ISBN: 0-387-13384-4] A report on the use of importance sampling (a Monte Carlo integration technique) for Bayesian analysis of simultaneous equation models. AO

**Applications (Engineering), P.** Annual Review of Fluid Mechanics, Volume 17, 1985. Ed: Milton Van Dyke, J.V. Wehausen, John L. Lumley. Annual Reviews, 1985, 630 pp, $28. [ISBN: 0-8243-0717-8]

**Applications (Engineering), T(14-15: 1), L.** Theory and Applications of Linear Differential and Difference Equations: A Systems Approach in Engineering. R.M. Johnson. Ser. in Math. & its Applic. Halsted Pr, 1984, viii + 183 pp, $29.95. [ISBN: 0-470-20106-1] Uses Laplace and z-transforms to study (via transfer functions, block diagrams, frequency domain analysis) the similarities between linear differential and difference equations. Examples from many fields of engineering. A final chapter studies digital filters. DFA

**Applications (Engineering), S(17), P.** Topics in Boundary Element Research, Volume 1: Basic Principles and Applications. Ed: C.A. Brebbia. Springer-Verlag, 1984, xiii + 256 pp, $49.50. [ISBN: 0-387-13097-7] Each of the ten chapters in this work describes a new application of boundary element methods. The chapters are self-contained and authored by different scientists who are actively involved in boundary element research. CEC

**Applications (Engineering), T(16-17: 4), P.** Mass Transfer: Fundamentals and Applications. Anthony L. Hines, Robert N. Maddox. Prentice-Hall, 1985, xiii + 542 pp, $37.95. [ISBN: 0-13-559609-2] Aimed at chemical engineering students. Good selection of problems in each chapter. First half on microscopic diffusional processes and prediction and use of transport coefficients. Second half on macroscopic separation processes, including absorption, distillation, and extraction. Useful reference with numerous graphs and tables and references at end of each chapter. BH

**Applications (Engineering), P.** Lecture Notes in Engineering-7: Boundary Integral Equation, Analyses of Singular, Potential, and Biharmonic Problems. Derek B. Ingham, Mark A. Kelmanson. Springer-Verlag, 1984, iv + 173 pp, $12.50 (P). [ISBN: 0-387-13646-0] Extensions of the boundary integral equation method to certain cases involving highly non-linear boundary conditions and governing elliptic equations. JAS

**Applications (Physics), P.** Mathematical Physics Reviews, Volume 4 (1984). Ed: S.P. Novikov. Transl: Morton Hamermesh. Harwood Academic, 1984, ix + 280 pp, $99.50. [ISBN: 3-7186-0146-X] English translations of four Russian papers in mathematical physics. Topics include the theory of "typical" singularities of functions and mappings, applications of Lie algebra methods to the classification of solutions of the Yang-Baxter equation, Hamiltonian formalism in the hydrodynamics of ideal fluids and the classification of integrable one-dimensional systems of certain types, which generalize the Korteweg-de Vries and Burgers equations. BH

**Applications (Physics), P.** An Approach to the Foundations of Quantum Mechanics, Using the Concept of A Filter as Primitive. Hans Kummer. Papers in Pure & Appl Math., No. 70. Queen's U, 1984, 160 pp, (P). Investigates use of the concept of an operation (or filter) as a primitive concept in an axiomatic theory. AO

**Applications (Physics), S(18), P.** Vertex Operators in Mathematics and Physics. Ed: J. Lepowsky, S. Mandelstam, I.M. Singer. Math. Sci. Res. Inst., V. 3. Springer-Verlag, 1985, xiv + 482 pp, $29.80. [ISBN: 0-387-96121-6] These proceedings of a 1983 research conference explore the connections between the affine Kac-Moody Lie algebras and the dual-string theory of particle physics through the use of formal differential operators called vertex operators. A poor attempt is made to provide an introduction suitable for the nonexpert reader. MU

**Applications (Physics), P.** Lecture Notes in Physics-211: Resonances--Models and Phenomena. Ed: S. Albeverio, L.S. Ferreira, L. Streit. Springer-Verlag, 1984, vi + 359 pp, $18 (P). [ISBN: 0-387-13880-3] Proceedings of a symposium held at the University of Bielefeld in July 1984. Papers present recent research on the modeling of resonance phenomena. AO

**Applications (Physics), S(18), P.** A Concept of Mathematical Physics: Models for Space-Time. Tamás Matolcsi. Akademiai Kiado, 1984, 236 pp, $26. [ISBN: 963-05-3245-X] Part one describes mathematical models of space-time, nonrelativistic and relativistic (both special and general) using an axiomatic approach. Part two presents the mathematical tools required for part one: tensors, pseudo-Euclidean spaces, affine spaces, smooth manifolds, and Lie groups. BH

**Applications (Physics), S(16-17), L.** Mathematical Analysis of Physical Problems. Philip R. Wallace. Dover, 1984, xix + 616 pp, $11.95 (P). [ISBN: 0-486-64676-9] Unabridged and corrected republication of the work originally published by Holt, Rinehart and Winston in 1972 (TR, November 1973). For the advanced undergraduate or beginning graduate student. From the vibrating string to the quantum mechanics of the many-body problem. Well-organized, useful as a reference, at a price that is hard to beat. Problems, but no solutions. JK

**Applications (Physics), P.** Quantum Theory of Particles and Fields: Birthday volume dedicated to Jan Łopuszański. Ed: Bernard Jancewicz, Jerzy Lukierski. World Scientific (US Dist: Heyden & Son), 1983, xxvi + 262 pp, $33. [ISBN: 9971-950-77-4] Collection of papers on the quantum theory of particles and fields. BH

**Applications (Physics), P.** Stochastic Methods and Computer Techniques in Quantum Dynamics. Ed: H. Mitter, L. Pittner. Springer-Verlag, 1984, vii + 452 pp, DM 122. [ISBN: 0-387-81835-9] Proceedings of the 23rd Internationale Universitätswochen für Kernphysik 1984 der Karl-Franzens-Universität Graz held at Schladming (Steiermark, Austria), February 20 to March 1, 1984. JAS

## Reviewers

DFA: David F. Appleyard, Carleton; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; RD: Roger Day, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; BH: Bruce Hanson, St. Olaf; PH: Paul Humke, St. Olaf; KK: Kenneth Kaminsky, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; AM: Alan Magnuson, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; LN: Linda Ness, Carleton; AO: Arnold Ostebee, St. Olaf; MS: Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MT: Michael Tveite, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

# Section Reports

An asterisk (*) by the title of a paper indicates that copies of the paper are available from the author. Papers presented under special sponsorship as part of joint meetings are so noted in parentheses.

## Pacific Northwest Section

The annual meeting of the Pacific Northwest Section was held at Williamette University, Salem, Oregon on June 20-22, 1985. Approximately 70 persons attended the meeting.

Invited Lectures:

"The Undergraduate Curriculum," by Donald W. Bushaw, Washington State University.
"Pictures and Stories from George Polya's Photo Album," by Gerald L. Alexanderson, University of Santa Clara.

"Graphical Data Analysis in More Than Two Dimensions," by Wesley L. Nicholson, Battelle-Northwest.

Short Presentations:

"On Certain Generalized Circular Transforms," by Stanley D. Luke, Seattle Pacific University.
"Multiplicative Magic Squares," by Ellis R. Von Eschen, Suffolk Community College.
"Chains in Finite Boolean Lattices," by Roger B. Nelson and Harvey Schmidt, Jr., Lewis and Clark College.
"Finite Calculus," by Rudolf Festa, Anchorage, Alaska.

Mini-Courses:

"Teaching Problem Solving," by Alan H. Schoenfeld, University of California, Berkeley.
"Geometry for College Teachers," by Branko Grunbaum, University of Washington.

Panel Discussion:

"Teaching Discrete Mathematics," by Willy Brandal, University of Idaho; Bryan Dorner, Pacific Lutheran University; Steve Prothero, Williamette University; Thomas Reimer, Lane Community College; Charles R.B. Wright, University of Oregon.

## Intermountain Section

The annual spring meeting of the Intermountain Section was held April 19-20, 1985 at Weber State College, Ogden, Utah. There were 81 persons in attendance.

Invited Addresses:

"The MAA Precollegiate Mathematics Contest," by Stephen B. Maurer, Swarthmore College.
"What is Discrete Algorithmic Mathematics?" by Stephen B. Maurer, Swarthmore College.

Short Presentations:

"The Family Nim," by Paul Yearout, Brigham Young University.
"Mathematical Problems Associated with NUSAT," by Paul Talaga, Weber State College.
"Mathematics Education in Russia," by Jim Cotts, Southern Utah State College.
"Permutations Without Invariant Elements," by Charles Cunkle, Emeritus, Slippery Rock.
"Pivot Point Graphing of Some Common Families of Functions," by John Thaeler, Weber State College.
"Bases for the Space of Invariants of a Class of Transformations," by L. Kirk Tolman, Brigham Young University.
"Romberg Integration for Unequal Spacing," by Donald R. Snow, Brigham Young University.
"How To Shoot Craps If You Must," by S.N. Ethier, University of Utah.
"Some Observations of Pairwise Votes," by Lawrence C. Ford, Idaho State University.
"Cryptography: An Applied Avenue to Topics in Pure Mathematics," by Lee Badger, Weber State College.
"The Bisection Method is Almost Never Linear," by Frank Richards, Weber State College.
"Fitting a Rational Function to a Finite Set of Points," by Jay Huber, Ricks College.

Panel Discussions:

"Certification in Mathematics," by Patricia Henry, Weber State College; Donald H. Tucker, University of Utah; Steven H. Heath, Southern Utah State College.
"Discrete Mathematics," by Lee Badger, Weber State College; Robert Girse, Idaho State University; LeRoy Beasley, Utah State University; John Higgins, Brigham Young University; Keith Reed, University of Utah.

Student Papers:

"A Model of a Problem in Genetics: Mutants Under Stress," by Kevin Black, Brigham Young University.
"A Model for Growth in Endothermie Animals," by Tim Thurston, Weber State College.
"A Model for the Growth of the Spruce Bud Worm," by Brian Heath, Weber State College.
"A Model for the Car Following Problem," by Pete Jahsman, Weber State College.
"Satellite Attitude Approximation Using Light Sensor Data," by Lance M. Moss, Weber State College.
"Simplifying JK Flip-Flops with the Boolean Algebra of Order Four," by Donald John Nicholson, University of Utah.
"Modeling Reserves for Group Health Plans," by Robert L. Bartholomew, Brigham Young University.
"Homomorphisms of Graphs and their Related Contractions," by Rick Gillman, Idaho State University.
"Applications of the Word Problem to Public Key Cryptography," by Jeff Salt, Weber State College.

# CENTER SECTION INDEX

## Volume 92, 1985

**Section Reports**

LEMMA. *Nondistinct solutions of* $S = \text{Circ}(1,1,1,0,\ldots,0)X$ *result* (i) *if two consecutive elements in* $S$ *are equal*; *or* (ii) *if* $S$ *contains five consecutive elements of the form abcba.*

*Proof.* The equation $x_j + x_{j+1} + x_{j+2} = x_{j+1} + x_{j+2} + x_{j+3}$ implies $x_j = x_{j+3}$, confirming (i). To confirm (ii), suppose, without loss of generality, that

$$x_1 + x_2 + x_3 = a,\ x_2 + x_3 + x_4 = b,\ x_4 + x_5 + x_6 = b,\ x_5 + x_6 + x_7 = a.$$

These equations imply $x_1 - x_4 = a - b = x_7 - x_4$, so $x_1 = x_7$.

The key question regarding the flim-flam game of Example (4) is: What can a player's secret number possibly be? Obviously it can't be, say, 14 because everyone has a 12 on their clockface and, next to it, at least 1 and 2. The bidding should begin at 16. The bids $17, 18, 19, 20, \ldots$ pass uneventfully.

Let $\sigma$ denote any secret number. To obtain a lower bound on $\sigma$ we compute $\bar{s}$, the average value of the triple-sums $s_j$:

$$\bar{s} = \frac{1}{12}\sum_{k=1}^{12} s_k = \frac{1}{4}\sum_{k=1}^{12} x_k = \frac{1}{4}\sum_{k=1}^{12} k = 19.5.$$

Thus, $\sigma \geqslant 20 > \bar{s}$. In actuality, $\sigma$ can never equal 20. Twelve integers, the largest of them 20, average out to 19.5 only if at least 6 of them equal 20. If exactly 6 of the $s_j$ equal 20, then the others must all be 19. In that case, either two consecutive $s_j$ are the same, or else the 20's and 19's alternate. Both possibilities are violations, in view of the above Lemma. If more than 6 of the $s_j$ equal 20, we have the same violation because (the pigeonhole principle!) at least two consecutive entries in $S$ must be 20.

Thus, $\sigma \geqslant 21$. Meanwhile, my accomplice is ready with a circular permutation for which $\sigma$ attains its lower bound. It is $1, 8, 10, 3, 5, 9, 4, 6, 11, 2, 7, 12$.

The only remaining question: What is the probability that someone in the audience has also stumbled onto a circular permutation of $1, 2, \ldots, 12$ for which $\sigma = 21$? Frankly, I don't know the probability, but I believe that it is very small. After all, my accomplice and I have never lost a game.

**Acknowledgment.** I wish to thank Professors P. J. Davis and O. Shisha for helpful discussions.

### References

**1.** C. Berge, Graphs and Hypergraphs, North-Holland, 1973.

**2.** E. J. Hoffman, J. C. Loessi, and R. C. Moore, Constructions for the solution of the *m* queens problem, Math. Mag., 42 (1969) 66–72.

**3.** G. Berman and K. D. Fryer, Introduction to Combinatorics, Academic Press, 1972.

**4.** D. I. A. Cohen, Combinatorial Theory, Wiley, 1978.

**5.** C. L. Liu, Elements of Discrete Mathematics, McGraw-Hill, 1977.

**6.** A. Tucker, Applied Combinatorics, Wiley, 1980.

## MULTIPLICATIVE RELATIONS FOR SUMS OF INITIAL $k$TH POWERS

DAVID E. PENNEY AND CARL POMERANCE*

*Department of Mathematics, University of Georgia, Athens, GA 30602*

Is the curious and pretty identity

$$(1 + 2 + \cdots + n)^2 = 1^3 + 2^3 + \cdots + n^3$$

*Supported in part by a National Science Foundation grant.

an anomaly or are there more such? Specifically, let $f_k$ denote the function whose value at the positive integer $n$ is

$$f_k(n) = 1^k + 2^k + \cdots + n^k.$$

We ask for a list of all identities of the form

$$\prod_{i=1}^{r} f_{h(i)}^{a(i)} = \prod_{j=1}^{s} f_{k(j)}^{b(j)}, \tag{1}$$

where the exponents are positive integers and the subscripts are distinct nonnegative integers. We shall show the only such identity is $f_1^2 = f_3$ and powers of this equation.

It is well known that $f_k$ is a polynomial of degree $k+1$. In fact,

$$f_k(n-1) = \frac{1}{k+1}\sum_{j=1}^{k+1}\binom{k+1}{j}B_{k+1-j}n^j,$$

where the $\{B_i\}$ are the Bernoulli numbers. A tidy way of defining these ubiquitous numbers is via the identity

$$\frac{x}{e^x-1} = \sum_{i=0}^{\infty}\frac{B_i x^i}{i!},$$

so that $B_i$ is the value of the $i$th derivative of $x/(e^x-1)$ at $x=0$. See [5], pp. 230–246, for proofs and a complete discussion.

Suppose that (1) holds. Without loss of generality we may assume that

$$h(1) < h(2) < \cdots < h(r) \quad \text{and} \quad k(1) < k(2) < \cdots < k(s);$$

moreover, that $h(r) < k(s)$. Let us first examine the case in which $k(s)$, the largest of the subscripts, is at most 3. Then the identity in (1) may involve only the polynomials

$$f_0(n) = n, \quad f_1(n) = \frac{1}{2}n(n+1),$$

$$f_2(n) = \frac{1}{6}n(n+1)(2n+1), \quad f_3(n) = \frac{1}{4}n^2(n+1)^2.$$

Because $f_2(n)$ has the irreducible factor $2n+1$ which appears in none of the other three polynomials, $f_2(n)$ is not involved in the identity. By examining the exponent on the factor $n+1$, we see that the only possible identity is $f_1^{2a} = f_3^a$ for some positive integer $a$.

Now suppose that $k(s) > 3$. By evaluating the polynomial identity in (1) at $n=2$ we obtain the integer equation

$$\prod_{i=1}^{r}[1+2^{h(i)}]^{a(i)} = \prod_{j=1}^{s}[1+2^{k(j)}]^{b(j)}. \tag{2}$$

From a theorem of Bang [2], there is a prime divisor $p$ of $1+2^{k(s)}$ such that 2 belongs to the exponent $2k(s)$ in the integers modulo $p$. Because $h(i) < k(s)$ for each $i$, we see that $p$ cannot be a divisor of the left-hand side in (2). Thus there can be no identity of the form in (1) with $k(s) > 3$. This completes the proof of our theorem.

The result of Bang to which we refer deals with "primitive" prime factors of expressions of the form $a^n - 1$ (we were concerned above with $2^{2k(s)} - 1$). This result was later generalized by Zsigmondy [9] to expressions of the form $a^n - b^n$:

ZSIGMONDY'S THEOREM. *If $a$, $b$, and $n$ are integers with $a > b > 0$, $\gcd(a,b) = 1$, and $n > 2$, then there is a prime divisor $p$ of $a^n - b^n$ such that $p$ is not a divisor of $a^k - b^k$ for any integer $k$ with $1 \leqslant k < n$, except for the case $a = 2$, $b = 1$, $n = 6$.*

Both Bang's theorem and Zsigmondy's theorem have been rediscovered many times in the last century. A partial list of references is given in [3], p. 361. It should also be noted that Zsigmondy's theorem has itself been generalized to algebraic number fields—a special case of this situation implies that Fibonacci numbers have primitive prime factors. A recent reference on generalizations of Zsigmondy's theorem is Stewart [7], from which earlier references may be tracked down.

Using Zsigmondy's theorem, we can handle the following generalization of (1). If $c, d \geqslant 1$, $k, n \geqslant 0$ are integers, let $f(c, d, k)$ be the function whose value at $n$ is

$$f(c,d,k;n) = \sum_{i=0}^{n-1} (c+di)^k. \tag{3}$$

Thus $f(1,1,k) = f_k$. For a fixed pair $c, d$ we can ask if there are any multiplicative identities

$$\prod_{i=1}^{r} f(c,d,h(i))^{a(i)} = \prod_{j=1}^{s} f(c,d,k(j))^{b(j)}, \tag{4}$$

where the $\{h(i)\}$ and $\{k(j)\}$ are distinct nonnegative integers. We leave it to the reader to verify the pleasant exercise that the only solutions of (4) are powers of the equations

$$f(1,1,1)^2 = f(1,1,3),$$

and

$$f(1,2,0)^2 = f(1,2,1).$$

Note that this result is a grand generalization of MONTHLY Problem E 2951 [**8**], which asked for a catalog of all equations

$$f(1,2,h)^a = f(1,2,k)^b,$$

and also the work of Edmonds [**4**] and Allison [**1**], who considered identities of the form

$$f(1,1,h)^a = f(1,1,k)^b.$$

This latter equation was also considered in MONTHLY Problem E 2136 [**6**].

We take this opportunity to thank Paul Bateman, Bruce Berndt, H. W. Gould, and William J. LeVeque for their assistance in locating references concerning the equation in (1). In light of the many rediscoveries of the Bang-Zsigmondy result, it would be ironic if our results here had been anticipated. In such an event, which is perhaps not unlikely, the responsibility is of course solely ours.

## References


**1.** D. Allison, A note on sums of powers of integers, this MONTHLY, 68 (1961) 272.

**2.** A. S. Bang, Taltheoretiske Undersøgelser, Tidsskrift Math., 5 IV (1886), 70–80 and 130–137.

**3.** G. G. Dandapat, J. L. Hunsucker, and C. Pomerance, Some new results on odd perfect numbers, Pacific J. Math., 57 (1975) 359–364.

**4.** S. M. Edmonds, Sums of powers of the natural numbers, Math. Gazette, 41 (1957) 187–188.

**5.** Charles Jordan (Jordân Kâroly), Calculus of Finite Differences, Chelsea, New York, 1965.

**6.** M. S. Klamkin, Solution to Problem E 2136, this MONTHLY, 76 (1969) 946–947 (proposed by A. Inselberg and B. Dimsdale, 75 (1968) 1113).

**7.** C. L. Stewart, On divisors of terms of linear recurrence sequences, J. Reine Angew. Math., 333 (1982) 12–31.

**8.** Nicholas Tzanakis, Solution to Problem E 2951, this MONTHLY, 91 (1984) 142–143 (proposed by Barbara Turner, 89 (1982) 273).

**9.** K. Zsigmondy, Zur Theorie der Potenzreste, Monatsh. Math. Phys., 3 (1892) 265–284.

Who is Louis de Branges? (See p. 742.)

Consequently $H(x)$ must be negative when $x$ is sufficiently large, and this is a contradiction. Hence $L = \frac{1}{4}$ and $\frac{1}{4}$ is not taken on by $b$.

Also solved by Nicholas Passell, Douglas B. Tyler, J. Vukmirović (Yugoslavia), and the proposer.

### On Areas of Triangles

6466 [1984, 441]. *Proposed by I. J. Schoenberg, University of Wisconsin-Madison, from a problem seminar of A. Ostrowski* (1924).

Let $F$ be a compact plane set, not contained in a line. Let

$$I = \sup\{\text{area of triangle } ABC\colon \{A, B, C\} \subset F\},$$

$$J = \inf\{\text{area of triangle } ABC\colon F \subset \text{closed triangle } ABC\}.$$

Show that $4I \geqslant J$ and that equality can occur.

*Composite Solution*. Since $F$ is compact there are points $A, B, C$ in $F$ such that $I =$ area of triangle $ABC$. Let $T$ be the closed triangle formed by the line $l$ through $A$ parallel to $BC$, the line $m$ through $B$ parallel to $AC$ and the line $n$ through $C$ parallel to $AB$. Then $T$ has area $4I$. Further $F \subset T$, for if $X$ is a point off $l$ (say) on the side away from $BC$, then triangle $XBC$ has area greater than $I$. Thus $J \leqslant 4I$.

Equality occurs when $F$ is a circle or when $F$ comprises the vertices of a square.

Solved by Miroslav D. Ašić (Yugoslavia), Richard Goldstein, John M. Ingram, Jeong-Han Kim, Sr. (Korea), N. J. Lord (England), A. McD. Mercer (Canada), O. P. Lossers (The Netherlands), Victor Pambuccian (Romania), Pier Ivan Pastro (Italy), Richard E. Pfiefer, John H. Riley, Jr., C. R. Selvaraj, University of South Alabama Problem Group, J. Vukmirović (Yugoslavia), and the proposer. Victor Pambuccian showed by an elaboration of the above argument that "compact" can be replaced by "bounded".

### A Counting Problem

6468 [1984, 441]. *Proposed by Calvin T. Long, Washington State University.*

If $a$ and $n$ are integers with $n > 1$, prove that $\sum_{d|n}\phi(d)a^{n/d} \equiv 0 \pmod{n}$. Note that if $p$ is a prime, this reduces to Fermat's Little Theorem, $a^p \equiv a \pmod{p}$.

*Solution*. Thirty-two solutions were submitted including one from the proposer. Several solvers pointed out that P. A. MacMahon [Proc. London Math. Soc., 23 (1891–2), 305–313] proved that

$$\frac{1}{n}\sum_{d|n}\phi(d)\,a^{n/d}$$

is the number of circular permutations of $n$ elements of $a$ kinds, with repetitions allowed. This solves the problem. It was also pointed out that the problem is essentially the same as problem E 2242, two solutions of which appear in this MONTHLY [1971, 545–546].

---

### ANSWER TO PHOTO ON PAGE 732

The photo is of Ludwig Bieberbach (1886–1982), the proposer of a famous conjecture bearing his name. The question is in continuation of a similar one that appeared in this MONTHLY, February 1985, page 93.

**References**

1. S. Sternberg, Lectures on Differential Geometry, Prentice-Hall, Englewood Cliffs, NJ, 1965.
2. R. Kaufman, A singular map of a cube onto a square, J. Differential Geom., vol. 14 (1979) 593–594.
3. H. Whitney, A function not constant on a connected set of critical points, Duke Math. J., vol. 1 (1935) 514–517.

---

# THE TEACHING OF MATHEMATICS

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

*Material for this department should be sent to Professor Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.*

## DISCRETE AND CONTINUOUS COMPOUNDING

MORRIS MORDUCHOW
*Department of Mechanical and Aerospace Engineering, Polytechnic Institute of New York, Brooklyn, NY 11201*

In [**1**] R. C. Thompson shows that if $g$ is the true growth rate in an inflationary economy, then

$$g = i - r \tag{1}$$

if the interest rate $i$ and the inflation rate $r$ are compounded continuously, whereas

$$g = \frac{i - r}{1 + r} \tag{2}$$

if the compounding is periodic. (In [**2**] Equation (2) is also presented, but in different terms.) There appears to be a contrast between the true growth rate in an inflationary economy, depending on whether the compounding is discrete or continuous. On the other hand, one expects that (2) approximates (1) if the compounding is performed over sufficiently small periods. The purpose of this note is to show how to modify (2) to take into account the *frequency* of compounding.

Consider a basic time-unit, say one year, and suppose that the interest rate $i$ is compounded "$(1/n)$thly", so that in one year an investment of $P$ dollars grows to

$$Q = P\left(1 + \frac{i}{n}\right)^n. \tag{3}$$

After $T$ years, the growth would be

$$Q = P\left(1 + \frac{i}{n}\right)^{nT}. \tag{4}$$

Similarly, at an inflation rate of $r$ compounded $(1/n)$thly an item costing $X$ dollars now would cost

$$Y = X\left(1 + \frac{r}{n}\right)^{nT} \tag{5}$$

after $T$ years. The value of $(Q/Y)$ at any time $T$ would denote the number of items that the investment could purchase at that time, and hence may be regarded as a measure of the purchasing power. The true growth rate $g$ may accordingly be defined by the growth of $(Q/Y)$, as follows:

$$\frac{Q}{Y} = \frac{P}{X}\left(1 + \frac{g}{n}\right)^{nT}. \tag{6}$$

Substituting for $Q$ and $Y$ from (4) and (5) into (6), and solving for $g$, one arrives at

$$g = \frac{i - r}{1 + \left(\frac{r}{n}\right)}. \tag{7}$$

Equation (7) gives the required modification of (2) to take into account the frequency of compounding. Note that (7) becomes (2) for $n = 1$, and approaches (1) as $n \to \infty$. The discussion of the "inflation balancing principle" of [1] (which remains valid for all $n$) can similarly be modified to take into account the frequency of the compounding.

### References


1. R. C. Thompson, The true growth rate and the inflation balancing principle, this MONTHLY, 3 (1983) 207–210.
2. Irving Fisher, Appreciation and interest, Publications of the American Economic Association, XI (August 1896) 331–442.


---

# PROBLEMS AND SOLUTIONS



*Send all **proposed** problems, typed and in duplicate if possible, to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given at the head of each problem set.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

## ELEMENTARY PROBLEMS

*Solutions of these Elementary Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by April 30, 1986. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgement).*

E 3117. *Proposed by William C. Waterhouse, Pennsylvania State University.*

The 1983 Putnam Competition asked entrants to show (in essence) that there exists a positive real number $u$ such that $[u^n]$ is odd for (positive) odd $n$ and even for even $n$. The published proof

[this MONTHLY, 91 (1984) 493] does not yield an explicit example. Find an explicit example of such a positive real number $u$.

E 3118. *Proposed by Gérard Letac, Université Paul Sabatier, France.*

Consider 3 positive independent random variables $U, X, Y$ such that $U$ is uniform on $(0,1)$ and $\max(UX, UY)$ is distributed like $X$.

(1) If the distribution of $Y$ is $\mu(dy) = (1+y)^{-2}\mathbf{1}_{(0,+\infty)}(y)\,dy$, show that $X$ has distribution $\mu$.

(2) If $X$ and $Y$ have the same distribution, show that there exists $c > 0$ such that $cX$ has distribution $\mu$.

E 3119. *Proposed by Paul K. Stockmeyer, College of William and Mary, and Editors.*

(a) Prove that for every integer $n \geqslant 3$ there exist sets $S_n \subseteq [0,1]$ such that $S_n$ contains no arithmetic progression of length $n$ and $S_n \cup \{x\}$ contains such a progression for every $x \in [0,1] - S_n$.

(b) For each $n \geqslant 3$ exhibit such a set $S_n$ by explicit construction.

E 3120. *Proposed by Kurt C. Foster and Lee A. Rubel, University of Illinois.*

Let $f$ and $g$ be two twice-continuously-differentiable functions on the interval $[0,1]$. Define

$$K_f(x)\left(=f''(x)\left[1+(f'(x))^2\right]^{-3/2}\right)$$

to be the curvature of the graph of $y = f(x)$ at the point $(x, f(x))$, with $K_g(x)$ similarly defined. Suppose $f(0) = g(0) = 0$ and $f'(0) = g'(0) = 0$, and that $K_g(x) \geqslant K_f(x)$ for all $x \in [0,1]$. Must $g(x) \geqslant f(x)$ for all $x \in [0,1]$?

E 3121. *Proposed by Calvin T. Long, Washington State University.*

Given any thirteen distinct real numbers, show that there exist at least two which satisfy the inequality

$$0 < \frac{x-y}{1+xy} < \sqrt{\frac{2-\sqrt{3}}{2+\sqrt{3}}}\,.$$

E 3122. *Proposed by Miha'ly Bencze, Brasov, Romania.*

Prove that

$$\sum_{k=1}^{n} \frac{1^k + 2^k + \cdots + n^k}{k \cdot n^k}\left(1 + (-1)^{k-1}\binom{n}{k}\right) = (n+1)\sum_{k=1}^{n} \frac{1}{k}.$$

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Irrational Series

E 2923 [1982, 63]. *Proposed by P. Erdős, Hungarian Academy of Sciences, and Claudia Spiro, University of Illinois.*

Let $1 < a_1 < a_2 < \cdots$ be an infinite sequence of integers. Prove that

$$\sum_{n=1}^{\infty} \frac{2^{a_n}}{a_n!}$$

is irrational.

*Solution by Ulrich Abel, Giessen, West Germany.* We obtain the following generalization:

$$S(r)=\sum_{n=1}^{\infty}\frac{r^{a_n}}{a_n!}$$

is irrational for every rational $r>0$. For the proof we assume $S(r)=b/c$ for $r=s/t$ with integers $b,c,s,t$. Then we choose a prime number

$$p>\max(s;t;c;e^r) \tag{1}$$

so that $c<a_n<p$ for some $n$. Let $N$ be the smallest integer with $a_{N+1}\geqslant p$. Then $p$ is a divisor of

$$\frac{a_{N+1}!}{a_N!} \tag{2}$$

and $a_N>c$. Therefore

$$a_N!S(r)=a_N!\left(\sum_{n=1}^{N}(a_n!)^{-1}r^{a_n}+R\right)$$

is an integer. By Taylor's formula, there exists a number $h, 0<h<1$, with

$$(a_{N+1}!)^{-1}r^{a_{N+1}}<R\leqslant\sum_{n=a_{N+1}}^{\infty}(n!)^{-1}r^n=(a_{N+1}!)^{-1}r^{a_{N+1}}e^{rh}.$$

Hence $R=(a_{N+1}!)^{-1}\,r^{a_{N+1}}\,e^x$, with $0<x<r$, and

$$t^{a_N}a_N!R=\frac{a_N!}{a_{N+1}!}\left(\frac{s}{t}\right)^{a_{N+1}}e^x t^{a_N}$$

is an integer. Thus, by (1) and (2), $p$ divides $e^x$. But this is a contradiction to $0<e^x<e^r<p$.

Also solved by R. Breusch, E. Butler, F. Dodd, G. Ehrlich, S. M. Gagola, Jr., M. Golomb, M. F. Kruelle, L. M. Levine, O. P. Lossers (The Netherlands), J. P. Robertson, J. M. Stark, K. L. Stellmacher, University of South Alabama Problem Group, and the proposers.

### Fractional Powers of Binomial Coefficients

E 2933 [1982, 212]. *Proposed by Doug Hensley, Texas A & M University.*

Show that for $1\leqslant k\leqslant n$

$$2\binom{n}{k}^{1/n}\geqslant\binom{n}{k-1}^{1/n}+\binom{n}{k+1}^{1/n}.$$

*Solution by V. D. Mascioni (student), Swiss Federal Institute of Technology, Zurich, Switzerland.* We will show that the inequality

$$\binom{n}{k}^{1/n}-\binom{n}{k-1}^{1/n}\geqslant\binom{n}{k+1}^{1/n}-\binom{n}{k}^{1/n}$$

is true for all $k$ such that $1\leqslant k\leqslant n$.

At first, we simplify both terms dividing or multiplying by the common factors and we get

$$\left(\frac{1}{k(n-k)}\right)^{1/n}-\left(\frac{1}{(n-k)(n-k+1)}\right)^{1/n}\geqslant\left(\frac{1}{k(k+1)}\right)^{1/n}-\left(\frac{1}{k(n-k)}\right)^{1/n},$$

or

$$\left(\frac{1}{(n-k)}\right)^{1/n}\left(\frac{1}{k^{1/n}}-\frac{1}{(n-k+1)^{1/n}}\right) \geqslant \left(\frac{1}{k}\right)^{1/n}\left(\frac{1}{(k+1)^{1/n}}-\frac{1}{(n-k)^{1/n}}\right).$$

Now, $f(k) = k^{-1/n} - (n-k+1)^{-1/n}$ is a strictly decreasing function of $k$ if $1 \leqslant k \leqslant n$. In fact,

$$f'(k) = -\frac{1}{n}\left(k^{-(n+1)/n} + (n-k+1)^{-(n+1)/n}\right) < 0$$

if $0 < k < n+1$. If $k \geqslant n/2$ we obtain also

$$\frac{1}{(n-k)^{1/n}} \cdot f(k) \geqslant \frac{1}{k^{1/n}} \cdot f(k) \geqslant \frac{1}{k^{1/n}} \cdot f(k+1),$$

and the inequality is verified for $k \geqslant n/2$. Using the relation $\binom{n}{k} = \binom{n}{n-k}$ and the fact that $k \geqslant n/2$ implies $n - k \leqslant n/2$, we immediately see that the inequality is true for all $k$ such that $1 \leqslant k \leqslant n$. Defining $\binom{n}{-1} = 0$, our result remains true for $k = 0$.

Also solved by R. Breusch, S. Heller, O. P. Lossers (The Netherlands), A. K. Lyzzaik (Saudi Arabia), L. E. Mattics, and the proposer.

## A Generalization of a Putnam Definite Integral

E 2960 [1982, 498]. *Proposed by Malcolm J. Sherman, State University of New York at Albany.*

Evaluate

$$\int_0^{\infty} \frac{dx}{(1+x^2)(1+x^{\alpha})}, \quad \text{where } \alpha \geqslant 0.$$

*Solution by Paul Smith, University of Victoria, Canada.* Setting $u = 1/x$ we obtain

$$\int_{1/t}^{t} \frac{dx}{(1+x^2)(1+x^{\alpha})} = \int_{1/t}^{t} \frac{u^{\alpha}du}{(1+u^2)(1+u^{\alpha})}.$$

Thus

$$\int_0^{\infty} \frac{dx}{(1+x^2)(1+x^{\alpha})} = \lim_{t\to\infty}\left(\frac{1}{2}\int_{1/t}^{t}\frac{dx}{1+x^2}\right) = \frac{\pi}{4}.$$

REMARKS

(1) The restriction $\alpha \geqslant 0$ is clearly unnecessary.

(2) The published solution [1] to Putnam problem A3 (1980) shows that $\int_0^{\pi/2} \frac{du}{1+\tan^{\alpha} u} = \frac{\pi}{4}$ for all $\alpha$. The substitution of $x = \tan u$ yields an alternate solution.

### Reference

1. L. F. Klosinski, G. L. Alexanderson, and A. P. Hillman, The William Lowell Putnam mathematical competition, this MONTHLY, 88 (1981) 605–612.

*Comment*: As pointed out by M. S. Klamkin (Canada), the problem appears in G. Pólya's *Induction and Analogy in Mathematics*, pp. 202, 275 with a generalization.

Also solved by 63 other readers and the proposer.

### The Limit Is Either 0 or $a^{1/c}$

E 2961 [1982, 498]. *Proposed by Ivan Niven, University of Oregon.*

Find all triples of positive constants $a, b, c$ such that

$$\lim_{n\to\infty}\left\{\frac{\sqrt[n]{a}+b}{c}\right\}^n$$

exists (as a finite limit), and determine its value.

*Solution by the proposer.* The limit exists if and only if $c \geqslant b+1$. If $c = b+1$ the limit is $a^{1/c}$. If $c > b+1$ the limit is zero. First assume $c = b+1$. Set aside the obvious case $a = 1$, and write

$$\left(\sqrt[n]{a}+b\right)/c = 1 + \left(\sqrt[n]{a}-1\right)/c = 1 + k_n/n,$$

by defining $k_n$ as $n(\sqrt[n]{a}-1)/c$. We note that

$$(*) \qquad (1+k_n/n)^n = \left[(1+k_n/n)^{n/k_n}\right]^{k_n}.$$

Also we find that $\lim k_n = (\log a)/c$ by applying L'Hôpital's rule to the quotient $(a^{1/x}-1)/(c/x)$. Hence

$$\lim k_n/n = 0 \quad \text{and} \quad \lim(1+k_n/n)^{n/k_n} = e.$$

Taking limits in $(*)$ we get

$$\lim\left[\left(\sqrt[n]{a}+b\right)/c\right]^n = \lim(1+k_n/n)^n = e^{(\log a)/c} = a^{1/c}.$$

If $c > b+1$, then, since $\lim\sqrt[n]{a} = 1$, there is a positive number $h$ such that $(\sqrt[n]{a}+b)/c < 1-h$ for all $n$ sufficiently large. Also, $\lim(1-h)^n = 0$. If $c < b+1$, there is a positive number $h$ such that $(\sqrt[n]{a}+b)/c > 1+h$ for all $n$ sufficiently large. And, of course, $(1+h)^n$ tends to infinity with $n$.

Also solved by 43 other readers, who all used essentially the same approach.

### Lattice Partitionings

E 2994 [1983, 287]. *Proposed by John E. Wetzel, University of Illinois.*

Let $Z^d$ be the set of lattice points in $\mathbb{R}^d$, the set of points all of whose coordinates are integers. For any subset $A$ of $Z^d$ write $\mathcal{D}(A)$ for the distance set of $A$, the set of distances $XY$ for $X, Y$ in $A$.

(a) Show that $Z^d$ can be partitioned into two subsets $A$ and $B$ so that neither $\mathcal{D}(A)$ nor $\mathcal{D}(B)$ equals $\mathcal{D}(Z^d)$.

(b) Show that if $A \cup B = Z^d$ but $A \cap B \neq \varnothing$, then either $\mathcal{D}(A)$ or $\mathcal{D}(B)$ must equal $\mathcal{D}(Z^d)$.

*Solution by Michael Josephy, University of Costa Rica, San José, Costa Rica.*

(a) Take

$$A = \left\{x = (x_1,\dots,x_d) \in Z^d \colon \textstyle\sum x_i \text{ even}\right\},$$
$$B = \left\{x = (x_1,\dots,x_d) \in Z^d \colon \textstyle\sum x_i \text{ odd}\right\}.$$

Then 1 belongs to neither $\mathcal{D}(A)$ nor $\mathcal{D}(B)$.

(b) If $A \cap B \neq \varnothing$, we can take (without loss of generality) $0 \in A \cap B$. Suppose $m = x_1^2$

$+ \cdots + x_d^2$, but $\sqrt{m} \notin \mathscr{D}(B)$. Then $x = (x_1, \ldots, x_d) \in A$.

Let us prove $\mathscr{D}(A) = \mathscr{D}(Z^d)$. If $\sqrt{n} \in \mathscr{D}(Z^d)$, then $n = y_1^2 + \cdots + y_d^2$. Let $y = (y_1, \ldots, y_d)$. If $y \in A$ or $x + y \in A$, then

$$\sqrt{n} = \|y - 0\| = \|(x + y) - x\|$$

is in $\mathscr{D}(A)$. Otherwise, $y \in B$ and $x + y \in B$, whence

$$\sqrt{m} = \|(x + y) - y\| \in \mathscr{D}(B),$$

a contradiction.

Also solved by C. Hurd, A. Kovačec (Austria), G. S. Lessells (Ireland), O. P. Lossers (The Netherlands), M. D. Meyerson, N. Miku (The Netherlands), D. Moews, Y. Peres and N. Kremerman (Israel), D. A. Rawsthorne, D. Richman, A. Rosenfeld, D. K. Skilton, and N. C. Wormald (Australia), H. G. Williams, and the proposer.

## ADVANCED PROBLEMS

*Solutions of these Advanced Problems should be mailed in duplicate to Professor D. H. Mugler, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053, by April 30, 1986. The solver's full post-office address should be on each sheet.*

6505. *Proposed by Harry Gonshor, Rutgers University.*

Prove that for any function $f\colon \mathbb{R} \to \mathbb{R}$ there is a function $g$ which differs from $f$ on a set of measure 0 and has the Darboux (i.e., intermediate value) property but is not continuous.

6506. *Proposed by Ronald Evans, University of California, San Diego.*

Let $p > 3$ be prime and write $f(n)$ for the Legendre symbol $(n/p)$. Prove that

$$\sum_{m=0}^{p-1} f(m^3 + 2)(1 - f(4m + 5)) = 1.$$

## SOLUTIONS OF ADVANCED PROBLEMS

### An Inequality for Rational Functions

6462 [1984, 371]. *Proposed by J. LeVeque and Lloyd N. Trefethen, Courant Institute of Mathematical Sciences, New York University.*

Let $S$ be the unit circle (or any other circle) in the complex plane. Let $r$ be a rational function of type $(n, n)$ with no poles on $S$.

(a) Show that $\|r'\|_1 \leqslant 4\pi n\|r\|_\infty$ where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are the $L_1$ and $L_\infty$ norms on $S$. Note that $\|r'\|_1$ is the arclength of the image of $S$ under $r$.

(b)* Show that $\|r'\|_1 \leqslant 2\pi n\|r\|_\infty$.

*Solution of part (a) by James C. Smith, University of South Alabama, Mobile, Alabama.* Take $S$ to be the unit circle and let $k = \max\{n, m\}$. We will show that, if $r$ is a rational function of type $(n, m)$ with no poles on $S$, then

$$\|r'\|_1 \leqslant 2(\pi k + n + m)\|r\|_\infty.$$

When $m = n$ the bound $2n(\pi + 2)\|r\|_\infty$ is better than $4\pi n\|r\|_\infty$.

We have $r = P/Q$ where $P$ and $Q$ are polynomials of degree $n$ and $m$, respectively. Let $f(\theta) = |r(e^{i\theta})|$ and $g(\theta) = \arg(r(e^{i\theta}))$. Then $f^2(\theta) = r(e^{i\theta})\overline{r(e^{i\theta})}$, where $\overline{r(e^{i\theta})}$ is a rational

function in $e^{-i\theta}$ of type $(n,m)$. It is easy to check that $e^{i(n-m)\theta}\frac{d}{d\theta}f^2$ can be expressed as $A/B$, where $A$ and $B$ are polynomials in $e^{i\theta}$ and the degree of $A$ does not exceed $2(n+m)$. Thus

$$\frac{d}{d\theta}f^2 = 2f\frac{d}{d\theta}f = 0$$

for at most $2(n+m)$ values of $\theta$ in $[0,2\pi)$. It follows that, for some $\alpha$, $[\alpha, \alpha+2\pi)$ is the union of at most $2(n+m)$ disjoint half-open intervals on which $f(\theta)$ is monotone and nonnegative, and hence that

$$\|f'\|_1 = \int_\alpha^{\alpha+2\pi}|f'| \leqslant 2(n+m)\|f\|_\infty .$$

Note that $r$ can be written as the product of linear fractional transformations say $T_1, T_2, \dots, T_k$. Since each $T_i$ maps $S$ one to one and onto a circle, we have

$$\int_0^{2\pi}\left|\frac{d}{d\theta}\arg(T_i)\right| \leqslant 2\pi.$$

Now, $g(\theta) = \Sigma\arg(T_i)$ and so $\|g'\|_1 \leqslant 2\pi k$.

In conclusion, since $r(e^{i\theta}) = f(\theta)e^{ig(\theta)}$, we have

$$r' = ie^{ig}fg' + e^{ig}f'.$$

Hence

$$\|r'\|_1 \leqslant \|fg'\|_1 + \|f'\|_1 \leqslant 2\pi k\|f\|_\infty + 2(n+m)\|f\|_\infty = 2n(\pi+2)\|r\|_\infty .$$

The proposers also solved part (a). Part (b)* remains open.

### A Least Upper Bound

6464 [1984, 440]. *Proposed by Roger Cooke, University of Vermont.*

What is the least upper bound of real numbers $b$ such that there exists a continuous real-valued function $f(x)$ satisfying

$$xf(x) - \int_1^x \left(f(t) + tf(t)^2\right)dt - b\ln(x) \to \infty \quad \text{as } x \to \infty?$$

*Solution by The University of South Alabama Problem Group, Mobile, Alabama.* Set $h(x) = xf(x) + (\frac{1}{2})$. Then we must find the least upper bound $L$ of real numbers $b$ satisfying

$$h(x) - \int_1^x \frac{h(t)^2}{t}\,dt - \left(b - \frac{1}{4}\right)\ln(x) \to \infty \quad \text{as } x \to \infty.$$

If $h(x) \equiv 0$ and $b < \frac{1}{4}$, then the limit is indeed infinite, so $L \geqslant \frac{1}{4}$. On the other hand we will show that $h(x) - \int_1^x \frac{h(t)^2}{t}\,dt$ takes on non-positive values for $x$ arbitrarily large.

Suppose, to the contrary, that there exists a continuous function $h(x)$ and an $x_0 \geqslant 1$ such that for $x \geqslant x_0$,

$$h(x) - \int_1^x \frac{h(t)^2}{t}\,dt > 0.$$

Set $H(x) = \int_1^x \frac{h(t)^2}{t}\,dt$, then for $x > x_0$, $h(x)^2 = xH'(x)$ so $\dfrac{H'(x)}{H(x)^2} > \dfrac{1}{x}$. Integrating both sides of this inequality from $x_0$ to $x$, we have

$$H(x_0)^{-1} - H(x)^{-1} > \ln(x) - \ln(x_0).$$

Consequently $H(x)$ must be negative when $x$ is sufficiently large, and this is a contradiction. Hence $L = \frac{1}{4}$ and $\frac{1}{4}$ is not taken on by $b$.

Also solved by Nicholas Passell, Douglas B. Tyler, J. Vukmirović (Yugoslavia), and the proposer.

### On Areas of Triangles

6466 [1984, 441]. *Proposed by I. J. Schoenberg, University of Wisconsin-Madison, from a problem seminar of A. Ostrowski* (1924).

Let $F$ be a compact plane set, not contained in a line. Let

$$I = \sup\{\text{area of triangle } ABC\colon \{A, B, C\} \subset F\},$$
$$J = \inf\{\text{area of triangle } ABC\colon F \subset \text{closed triangle } ABC\}.$$

Show that $4I \geqslant J$ and that equality can occur.

*Composite Solution.* Since $F$ is compact there are points $A, B, C$ in $F$ such that $I$ = area of triangle $ABC$. Let $T$ be the closed triangle formed by the line $l$ through $A$ parallel to $BC$, the line $m$ through $B$ parallel to $AC$ and the line $n$ through $C$ parallel to $AB$. Then $T$ has area $4I$. Further $F \subset T$, for if $X$ is a point off $l$ (say) on the side away from $BC$, then triangle $XBC$ has area greater than $I$. Thus $J \leqslant 4I$.

Equality occurs when $F$ is a circle or when $F$ comprises the vertices of a square.

Solved by Miroslav D. Ašić (Yugoslavia), Richard Goldstein, John M. Ingram, Jeong-Han Kim, Sr. (Korea), N. J. Lord (England), A. McD. Mercer (Canada), O. P. Lossers (The Netherlands), Victor Pambuccian (Romania), Pier Ivan Pastro (Italy), Richard E. Pfiefer, John H. Riley, Jr., C. R. Selvaraj, University of South Alabama Problem Group, J. Vukmirović (Yugoslavia), and the proposer. Victor Pambuccian showed by an elaboration of the above argument that "compact" can be replaced by "bounded".

### A Counting Problem

6468 [1984, 441]. *Proposed by Calvin T. Long, Washington State University.*

If $a$ and $n$ are integers with $n > 1$, prove that $\sum_{d|n}\phi(d)a^{n/d} \equiv 0 \pmod{n}$. Note that if $p$ is a prime, this reduces to Fermat's Little Theorem, $a^p \equiv a \pmod{p}$.

*Solution.* Thirty-two solutions were submitted including one from the proposer. Several solvers pointed out that P. A. MacMahon [Proc. London Math. Soc., 23 (1891–2), 305–313] proved that

$$\frac{1}{n}\sum_{d|n}\phi(d)\,a^{n/d}$$

is the number of circular permutations of $n$ elements of $a$ kinds, with repetitions allowed. This solves the problem. It was also pointed out that the problem is essentially the same as problem E 2242, two solutions of which appear in this MONTHLY [1971, 545–546].

---

### ANSWER TO PHOTO ON PAGE 732

The photo is of Ludwig Bieberbach (1886–1982), the proposer of a famous conjecture bearing his name. The question is in continuation of a similar one that appeared in this MONTHLY, February 1985, page 93.

EDITED BY ALLAN L. EDMONDS AND JOHN H. EWING
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER

*Basic Algebra*. By Nathan Jacobson. Volume I (1974) and II (1980). W. H. Freeman, San Francisco. *Algebra* (second edition). By Saunders Mac Lane and Garrett Birkhoff. Macmillan Publishing Company, New York, 1979.

KENNETH BOGART
*Department of Mathematics, Dartmouth College, Hanover, NH 03755*

It is ironic that two of the most modern texts in abstract algebra, two of the best examples of the power of abstraction, have neither the word "modern" nor the word "abstract" in their titles. This modesty signals the beginning of a new style in abstract algebra texts. More than a decade ago, this reviewer lamented the difficulty of teaching to undergraduates from the first edition of *Algebra*, a book which was the first to depart significantly from the Van der Waerden tradition. The second edition maintains this departure but as a result of reorganization is much more accessible to undergraduates. The first volume of Jacobson's *Basic Algebra* (BAI) appears to be much more in the Van der Waerden tradition, but has been written so that the nontraditional beginning of Volume II (BAII) is no surprise.

What are the ingredients that mark the departure from tradition? The most striking change is the early use of universal mapping properties in characterizing algebraic objects. Informally, a universal characterization of an object describes a property of the object and then asserts that anything else with the property is related by a morphism to the original object. Typically, the property can be expressed in terms of morphisms as well. Thus, universality is an idea of category theory. (In much the same way that group theory abstracts the study of the automorphisms of a single mathematical object, category theory abstracts the study of homomorphisms among mathematical objects.) As we learn in *Algebra*, virtually all the familiar algebraic constructions—rings of quotients, homomorphic images, direct products, even the natural number 1—are universal constructions. Chapter 4 of *Algebra* explains universality in terms of functors to the category of sets. When they are introduced in Chapter 4, categories and functors appear just as natural a part of algebra (and *Algebra*) as groups and homomorphisms.

Jacobson uses universal characterizations to good effect throughout BAI without stopping to describe how they are all manifestations of the same idea. In Chapter 1 of BAII, he gives us a discussion of category theory that expresses universality in terms of functors between arbitrary categories. The uninitiated reader might not recognize the extent to which universality pervaded BAI from the discussion in BAII, but the reader astute enough to ask the "obvious" questions will have the reward of being able to answer them. Later chapters of BAII—especially those on modules and homological algebra—make excellent use of the categorical underpinnings from Chapter 1.

A second departure from tradition of these texts is the introduction of ideas from universal algebra. In universal algebra, we study algebraic systems (sets with operations defined on them) in general rather than concentrating on a particular system. We study concepts such as homomorphisms, product constructions, subalgebras, etc., that can apply to all these systems. Further, we study families of algebras satisfying certain axioms; if these happen to be the axioms of a familiar system such as a group, then we are studying properties that distinguish that system from others. In *Algebra* the principal ideas from universal algebra are the almost traditional lattice theoretic ideas. Lattices are used primarily as an interesting algebraic system rather than being applied to the study of other algebraic systems. In BAII, however, Chapter II is almost a course in universal algebra, covering the basic isomorphism theorems, direct and inverse limit constructions, free algebras and varieties of algebras. These topics do not play a heavy role in the remainder of the

volume; this is more a testimony on the lack of influence of universal algebra on recent developments in "core" algebra than it is a criticism of the text. What is unfortunate is that neither text adopts from universal algebra the appealing idea of the equivalence of morphisms and congruence relations—a congruence relation for an operation like group composition preserves that operation in the same way that congruence modulo $n$ preserves addition and multiplication. This idea ties together all the basic homomorphism theorems of groups, rings, vector spaces, etc., as manifestations of a common idea, and, for example, makes it natural that only normal subgroups appear in the homomorphism theory of groups while all subspaces appear in the homomorphism theory of vector spaces (and no sublattice of a lattice plays the role of the kernel of a homomorphism). The fact that virtually all our familiar algebraic systems are special cases of monoids is clear in both books; this point of view is certainly in harmony with universal algebra.

In both texts, ideas involving linearity are discussed for modules over rings rather than vector spaces over fields. This aids our real understanding of these ideas by showing what properties of the rings are actually used at each stage of development. Even better, ideas from different areas of the traditional organization appear as corollaries of the same idea. For example, in *Algebra* determinants and tensor products live together and appear later with exterior algebras (an even more natural marriage than determinants and tensors). In BAI, determinants appear with exterior algebras. In both, the theories of finite abelian groups and the rational canonical form cohabitate.

*Algebra* follows one tradition from the Birkhoff and Mac Lane "Survey Series", namely linear algebra's role is more predominant than in other texts. Perhaps 40% of the book consists of topics we would normally associate with linear and multilinear algebra. This book is well suited to give all the *advanced* training in algebra an undergraduate needs. In fact, as one who teaches graduate courses in combinatorics using topics from widely separated branches of algebra, I would be delighted if my graduate students had mastered the concepts in *Algebra*.

In its two volumes, *Basic Algebra* comprises an entire education, advanced undergraduate and graduate, in algebra, an education deep enough to serve as a springboard to research in most traditional areas of algebra. It is unlikely that many graduate students, even those specializing in algebra, have learned such a breadth of ideas in such depth. To my knowledge, only Bourbaki has attempted as deep a survey of algebra as appears in BAII. (Of course in their encyclopedic way, the Bourbaki group covers even more.) Though linear and multilinear algebra is woven throughout BAI and BAII, it does not play as prominent a role in *Basic Algebra* as in *Algebra*. With that exception, BAII takes us on a tour of "core" algebra including commutative algebra, the structure theory of rings and algebras, representation theory for groups, and field theory in addition to its category theory, universal algebra and homological algebra. (Homological algebra, which arose from work on chains and cochains in algebraic topology, is based on the study of sequences of modules connected by homorphisms. Homological algebra has provided proofs of results of algebraic interest, some appearing in BAII, as well as results of topological interest. In the guise of exact sequences, it has had an influence on *Algebra* as well.)

Of course, it is unfair to expect a textbook to contain a complete compendium of results; however, I admit to disappointment that after developing the machinery needed to prove the unique factorization theorem for regular local rings, BAII does not include it even as a series of exercises. To a lesser extent, I felt a similar unsatisfied feeling when the results on modular and distributive lattices were not used in the appropriate places in ring theory (for example, a Dedekind domain is a domain whose lattice of ideals is distributive, and the modularity of the lattice of submodules of a module has not been exploited). When defining a coalgebra, Professor Jacobson comments that "It is nice to have a pretty definition, but it is even nicer to have some pretty examples." His examples are undeniably pretty, but in neither one do we learn what makes the coalgebra structure of interest. Frequently, the examples leave something to our imagination in this way. Since there is a tradeoff between the amount covered and the details included in a book, we should not regard the rather telegraphic style of examples in BAII as a defect. However, even the experienced mathematician can profit from discussions of BAII with a more knowledgeable colleague.

What role can these books play in a college mathematics curriculum? Both *Algebra* and BAI are appropriate for a one or two semester course in abstract algebra. If this were an undergraduate course, I'd want to restrict enrollment to "honors" students, successful mathematics students who are motivated by a genuine interest in mathematics. I would also be much more comfortable if the students had had a linear algebra course—a highly geometric version of the sophomore level course offered by many institutions would be totally appropriate, and would make the student's relationship with *Algebra* especially rewarding. With a strict selection of topics (omitting entire chapters), *Basic Algebra* I seems ideally suited for a two semester or three quarter course which would cover all the standard topics (including Galois theory) in the groups-rings-fields-vector spaces tradition. However, separate study in linear algebra would be essential to fill out the traditional program. *Algebra* would shine in a three semester sequence. Again, the sequence would cover all the traditional topics (Galois theory *is* in the second edition of *Algebra* and was not in the first) but would also provide an advanced linear and multilinear algebra background and touch gently on categories. As undergraduate texts, both lack a sufficient variety of straightforward exercises to give students practice with routine techniques. Both also expect the reader to think about the examples rather than merely digest them. At times the attention to detail in *Basic Algebra* obscures the idea behind a proof. Thus an instructor using one of these books will still have some work to do.

For graduate students, one might use portions of BAI in a one semester course or use *Algebra* in a two semester course. Either of these courses could be profitably followed by a course from BAII. How long would the course last? As long as possible! Realistically, there are at least four semesters worth of work in BAII. This volume is Professor Jacobson's gift to those of us who use algebra; I believe that as many mathematicians as possible should be exposed to as much of it as possible.

The authors of two of the most successful traditional series in abstract algebra have given us books with a fresh and much more modern perspective in algebra. We know now how algebra should be taught in the future.

(The reviewer would like to thank Ernst Snapper for valuable conversations about this review.)

---

*The Analysis of Linear Partial Differential Operators, Vols. I & II.* By Lars Hörmander. Grundlehren der mathematischen Wissenschaften 256 & 257, Springer-Verlag, Berlin-Heidelberg, 1983. ix + 391 pp.

MICHAEL E. TAYLOR
*Department of Mathematics, SUNY at Stony Brook, Stony Brook, NY 11794*

**History.** Calculus was perfected by Newton to formulate and solve differential equations arising in mechanics. Work on continua, modeling things such as moving fluids and vibrating solids, quickly led to the formulation of partial differential equations, and the subject of PDE has enjoyed a long and productive history. A number of important developments in analysis have been motivated by, and particularly effective in, the development of linear PDE, the subject of the volumes by Professor Hörmander.

What makes linear equations special is that the superposition principle applies:

$$Pu = f, \qquad Pv = g \Rightarrow P(u + \alpha v) = f + \alpha g. \tag{1}$$

It was noticed early that numerous constant coefficient linear PDE that arose naturally had special solutions in terms of trigonometric functions, or exponentials, seen to be equivalent in view of Euler's formula

$$e^{ix} = \cos x + i \sin x. \tag{2}$$

Thus it was tempting to exploit the superposition principle by representing general functions as superpositions of trigonometric functions. This idea was proposed by Daniel Bernoulli, as a tool for solving the wave equation

$$(3) \qquad (\partial^2/\partial t^2 - \partial^2/\partial x^2)u = 0$$

arising in the study of vibrating strings (in the linear approximation). The idea was not accepted at that time, but it began to gain reluctant acceptance after being reintroduced by Fourier as a tool in his investigation of the heat equation

$$(4) \qquad (\partial/\partial t - \partial^2/\partial x^2)u = 0.$$

Today, the Fourier transform, defined for integrable functions on $R^n$ by

$$(5) \qquad \mathscr{F}f(\xi) = \hat{f}(\xi) = (2\pi)^{-n/2}\int f(x)e^{-ix\cdot\xi}\,dx,$$

is a cornerstone of analysis. The Fourier inversion formula returns $f(x)$ from $\hat{f}$:

$$(6) \qquad f(x) = (2\pi)^{-n/2}\int \hat{f}(\xi)e^{ix\cdot\xi}\,d\xi.$$

The Fourier transform intertwines differentiation and multiplication:

$$(7) \qquad \mathscr{F}(\partial/\partial x_j) = i\xi_j\mathscr{F}, \qquad \mathscr{F}x_j = i(\partial/\partial\xi_j)\mathscr{F},$$

and hence it is often effective in reducing constant coefficient PDE to algebraic problems, or to ODE with parameters. Note in particular that

$$(8) \qquad \Delta = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_n^2 \Rightarrow \mathscr{F}\Delta = -|\xi|^2\mathscr{F};$$

$\Delta$ is called the Laplace operator. The wave and heat equations in $n$ space variables are, respectively,

$$(9) \qquad (\partial^2/\partial t^2 - \Delta)u = 0,$$

and

$$(10) \qquad (\partial/\partial t - \Delta)u = 0.$$

Natural exploitation of the Fourier transform led people to the limits of understanding the integral as it was understood in the eighteenth and nineteenth centuries, and the production of natural classes of functions (or other objects) on which the Fourier transform operates was achieved only in this century. A milestone was Lebesgue's solution of the problem of producing the right notion of integral of functions. Out of this came the Banach spaces $L^p(R^n)$, $1 \leqslant p \leqslant \infty$, and particularly the Hilbert space $L^2(R^n)$, which played a major part in the theory of functional analysis developing around the turn of the century. One aspect of the Fourier inversion formula is that $\mathscr{F}$ is a unitary operator on $L^2(R^n)$.

The development of theories of bounded operators on Banach and Hilbert spaces afforded only indirect attacks on partial differential equations, through the use of integral equations. One extension of this was the study of unbounded operators on Hilbert space, pursued by von Neumann, and later by K. O. Friedrichs and many others, which has been very influential. But the major advance was surely the creation of the theory of distributions.

**Distributions.** Distributions such as the Dirac delta function $\delta(x)$, defined by

$$(11) \qquad \int f(x)\delta(x)\,dx = f(0),$$

had arisen as a convenient language for describing certain "ideal" solutions to PDE. It was realized that, in some sense, $\delta(x)$ was the derivative of the Heaviside function $H(x)$, defined to be 1 for $x \geqslant 0$, 0 for $x < 0$. L. Schwartz showed in a very elegant fashion that the methods of

functional analysis could be applied to the study of such "generalized functions," to produce topological vector spaces (particularly $\mathscr{D}'(R^n)$, $\mathscr{E}'(R^n)$, and $\mathscr{S}'(R^n)$ as described below) containing all functions in $L^1(R^n)$ (compactly supported in the case of $\mathscr{E}'(R^n)$), and also closed under the action of differentiation. The Schwartz kernel theorem provided the basis of treating general operators as being generalized integral operators.

Also, the neatest presentation of Fourier analysis on $R^n$ was seen to exploit the Fourier transform on the Schwartz space $\mathscr{S}(R^n)$, defined by

$$u \in \mathscr{S}(R^n) \Leftrightarrow x^\beta D^\alpha u \in L^\infty(R^n) \quad \text{for all } \alpha, \beta. \tag{12}$$

Here $x^\beta = x_1^{\beta_1} \cdots x_n^{\beta_n}$, and $D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n}$, with $D_j = i\partial/\partial x_j$. One has $\mathscr{F}\colon \mathscr{S}(R^n) \to \mathscr{S}(R^n)$, and the Fourier inversion formula implies $\mathscr{F}$ is an isomorphism. By duality there is an isomorphism

$$\mathscr{F}\colon \mathscr{S}'(R^n) \to \mathscr{S}'(R^n); \tag{13}$$

the space $\mathscr{S}'(R^n)$ of tempered distributions is the space of continuous linear functionals on $\mathscr{S}(R^n)$, i.e., its dual space. There are natural inclusions

$$\mathscr{S}(R^n) \subset L^2(R^n) \subset \mathscr{S}'(R^n), \tag{14}$$

induced by

$$\langle u, v \rangle = \int_{R^n} u(x) v(x)\, dx. \tag{15}$$

The space $\mathscr{D}'(R^n)$ is dual to $C_0^\infty(R^n)$, the space of smooth, compactly supported functions, and the space $\mathscr{E}'(R^n)$ is dual to $C^\infty(R^n)$.

**Distributions and PDE .** With the machinery of distributions, it is convenient to formulate and exploit the notion of fundamental solution to a linear PDE:

$$PE_y = \delta(x - y). \tag{16}$$

If $P$ has constant coefficients, one can take $E_y(x) = E(x - y)$. In general, $E$ will be a distribution. One has a similar notion of fundamental solution of an initial value problem. For example, the fundamental solution to the heat equation would satisfy (10) for $t > 0$ and also

$$\lim_{t \downarrow 0} E(t, x) = \delta(x). \tag{17}$$

Let us indicate how Fourier analysis produces this fundamental solution. If $\hat{E}(t, \xi)$ denotes the partial Fourier transform with respect to $x$, the PDE (10) leads to

$$(d/dt)\hat{E}(t, \xi) = -|\xi|^2 \hat{E}(t, \xi), \tag{18}$$

and (17) yields

$$\hat{E}(0, \xi) = (2\pi)^{-n/2}. \tag{19}$$

Hence

$$\hat{E}(t, \xi) = (2\pi)^{-n/2} e^{-t|\xi|^2}. \tag{20}$$

Calculating $E(t, x)$ from (20) is equivalent to computing the Fourier transform of a Gaussian function, a basic computation in the subject; one obtains

$$E(t, x) = (4\pi t)^{-n/2} e^{-|x|^2/4t}. \tag{21}$$

Similarly, the Poisson kernel $P(y, x)$, satisfying

$$\begin{gathered}(\partial^2/\partial y^2 + \Delta) P(y, x) = 0, \qquad \text{for } y > 0,\ x \in R^n, \\ \lim_{y \downarrow 0} P(y, x) = \delta(x),\end{gathered} \tag{22}$$

is characterized by its partial Fourier transform with respect to $x$:

$$\hat{P}(y,\xi) = (2\pi)^{-n/2}e^{-y|\xi|}. \tag{23}$$

The computation of the inverse Fourier transform of this is easy if $n = 1$, but not so simple if $n > 1$. One way to obtain a formula for $P(y,x)$ is to exploit the "subordination identity":

$$e^{-y|\xi|} = \frac{1}{2}y\pi^{1/2}\int_0^\infty e^{-y^2/4t}e^{-t|\xi|^2}t^{-3/2}\,dt. \tag{24}$$

Application to (23), together with (21), gives

$$P(y,x) = c_n y(y^2 + |x|^2)^{-(n+1)/2}. \tag{25}$$

In order to prove (24), one might note it suffices to obtain an independent proof of its implication (25), for any one special case. As the direct verification that (23) implies (25) for $n = 1$ is straightforward, one has a proof of (24). The formula (25) for $P(y,x)$ can be analytically continued in $y$ for $\operatorname{Re} y > 0$, and one can pass to the limit of purely imaginary $y$, to obtain formulas for the fundamental solution to the wave equation:

$$(\partial^2/\partial t^2 - \Delta)W = 0, \qquad W(0,x) = 0, \qquad (\partial/\partial t)W(0,x) = \delta(x), \tag{26}$$

which satisfies

$$\hat{W}(t,\xi) = |\xi|^{-1}\sin t|\xi|. \tag{27}$$

If we note that the inverse Fourier transform of $(2\pi)^{-n/2}|\xi|^{-1}e^{-y|\xi|}$ is obtained by integrating (25) from $y$ to $\infty$, so is equal to

$$c_n'(y^2 + |x|^2)^{-(n-1)/2}, \tag{28}$$

for $n \geqslant 2$, it follows that

$$W(t,x) = \lim_{\varepsilon\downarrow 0} c_n' \operatorname{Im}\left(|x|^2 - (t - i\varepsilon)^2\right)^{-(n-1)/2}. \tag{29}$$

In particular, for $n = 3$, one obtains the formula

$$W(t,x) = (4\pi t)^{-1}\delta(|x| - t). \tag{30}$$

Integral formulas equivalent to these formulas for fundamental solutions are classical, of course, but the calculations (18)–(30) might give a feel for the utility and power of distribution theory and Fourier analysis.

A precise study of constant coefficient equations often enables one to get a hold on variable coefficient PDE, treating $P(x,D) = \sum_{|\alpha|\leqslant m} a_\alpha(x)D^\alpha$ as a perturbation of a constant coefficient operator obtained by freezing the coefficients. For example, if $P(x,D)$ is elliptic, which means

$$P_m(x,\xi) = \sum_{|\alpha|=m} a_\alpha(x)\xi^\alpha \neq 0 \qquad \text{for } \xi \neq 0, \tag{31}$$

then fundamental solutions for the constant coefficient elliptic operators can be pasted together to give the leading term in an inverse of $P(x,D)$. An iterative process will improve this approximation. For hyperbolic equations, such as the wave equation (9), in the variable coefficient case, more powerful methods are required. The volumes under review concentrate on the study of constant coefficient operators and some results on variable coefficient operators which one can deduce by such perturbations. More powerful methods, involving particularly pseudodifferential operators and Fourier integral operators, will be developed in following volumes by the author and applied to large classes of variable coefficient PDE.

**The Book.** The first of Hörmander's volumes is devoted primarily to the theory of distributions. It starts on a very basic level, with a detailed study of calculus of $C^\infty$ functions of several

variables, and proceeds with masterly organization to develop the theory to an advanced level. Specific fundamental solutions such as mentioned above are constructed. Actually, the author develops the material quite far before introducing the Fourier transform. Homogeneous distributions are introduced, and the behavior of $\Delta$ on $R^n$ under dilations leads us to look for a homogeneous distribution as a fundamental solution (at least for $n \geqslant 3$); rotational symmetry settles what it has to be, up to a scalar multiple. Change of variables enables one to treat any second order constant coefficient elliptic operator $\Sigma a_{jk} D_j D_k$, with $\Sigma a_{jk} \xi_j \xi_k$ positive definite. Analytic continuation is used to pass to fundamental solutions of second order hyperbolic operators. When the Fourier transform is introduced, some of these formulas are rederived, though by that point the emphasis is more on general classes of constant coefficient PDE, where analytical identities are not so available. One has for example the construction of a fundamental solution for an arbitrary constant coefficient operator $P(D)$. Use of Fourier analysis leads to studying $P(\xi)^{-1}$, for $\xi$ in the complex domain, and one needs a detailed understanding of the behavior of polynomials near their zero sets.

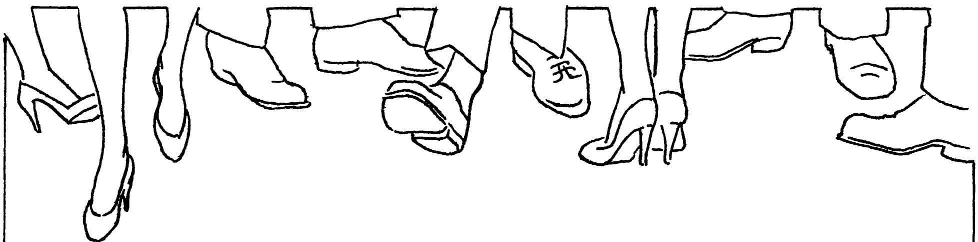
In studying solutions to a PDE $Pu = f$, one thing that is important is a description of the singularities of the solution. In particular, one wants to describe the singularities of the fundamental solution. For the wave equation in $n$ space variables, formula (29) shows the fundamental solution is singular on the cone $|x| = |t|$. A refinement of the notion of singular support has been introduced, the wave front set. Given a distribution $u \in \mathscr{D}'(R^n)$, $WFu$ is a subset of the cotangent bundle $T^*R^n$, projecting onto the singular support of $u$ in $R^n$, telling not only where $u$ is singular, i.e., where one cannot introduce a cut-off $\psi$ so that $\psi u \in C_0^\infty$, but also where the Fourier transform of $\psi u$ fails to tend rapidly to zero. In the case of a distribution with a simple singularity along a smooth surface, $WFu$ typically consists of all nonzero cotangent vectors normal to the surface. It is shown that for elliptic $P$, even with variable coefficients, $WF(Pu) = WFu$. This contains the elliptic regularity theorem. Volume I also considers notions of wave front set associated to other categories than $C^\infty$, for example the analytic wave front set, and ends with an introduction to hyperfunctions, an analogue of distributions, consisting of locally finite sums of analytic functionals. The theory of hyperfunctions has been developed quite far by M. Sato and collaborators, making heavy use of homological algebra. The introduction given here simplifies some of the foundations of that subject.

With the basic theory of distributions and Fourier analysis thoroughly set down, Volume II proceeds deeper into the study of linear PDE, particularly with constant coefficients. Detailed analysis is made of the regularity of certain fundamental solutions of general constant coefficient operators $P(D)$, including a classification of when $P(D)$ is hypoelliptic, i.e., satisfies

$$\operatorname{sing\,supp} Pu = \operatorname{sing\,supp} u, \tag{32}$$

or even $WF(Pu) = WFu$. A special class of hypoelliptic operators is the class of elliptic operators; the heat operator is also hypoelliptic. The wave front set of a fundamental solution $E$ of $P(D)$ is studied in great detail. Also questions of global solvability of an equation $P(D)u = f$ on a domain $\Omega \subset R^n$ are studied. Topics on variable coefficients include operators with "constant strength," and also a chapter on scattering theory, for short range potentials. As mentioned, subsequent volumes will go more deeply into the study of variable coefficient linear PDE.

These volumes are to some degree a greatly extended rewrite of Hörmander's *Linear Partial Differential Operators*, published by Springer-Verlag in 1964. They cover an enormous amount of analysis. In addition to topics mentioned above, there are self-contained treatments of topics ranging from the Gauss-Green formula and Cauchy's integral formula, to a complete though brief discussion of manifolds, differential forms, and Hamiltonian vector fields, to the Malgrange preparation theorem and the Tarski-Seidenberg theorem. Some of these topics will be more completely appreciated when they are used in subsequent volumes. The two volumes which are out, and their companions which will follow, will not likely serve as the texts for one's first brush with PDE, but the serious analyst will find here an elegant presentation of a vast amount of material on linear PDE, by a consummate master of the subject.



# The Mathematical Association of America
## CORPORATE MEMBERS

*The Mathematical Association of America gratefully acknowledges the interest and support of the following:*

CORPORATE MEMBERS

**Addison-Wesley Publishing Company, Inc.**
**American Mathematical Society, Inc.**
**Birkhäuser Boston, Inc.**
**Digital Equipment Corporation**
**International Business Machines Corporation**
**Macmillan Publishing Company**
**Marcel Dekker, Inc.**
**The Mitre Corporation**
**Pergamon Press, Inc.**
**Prindle Weber and Schmidt, Inc.**
**Saunders College Publishing, Inc.**
**Scott Foresman & Company, Inc.**
**Springer-Verlag New York, Inc.**
**John Wiley & Sons, Inc.**

CONTRIBUTING CORPORATE MEMBER

**Bell Laboratories, Inc.**

*Carus Mathematical Monograph #21*

## From Error Correcting Codes through Sphere Packings to Simple Groups,

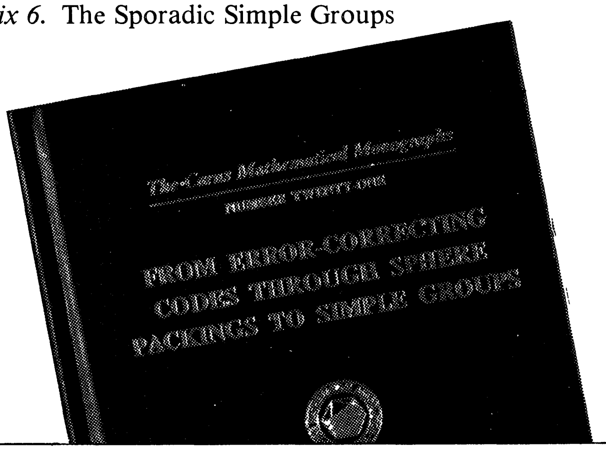
by Thomas M. Thompson
224 pp. Hardbound
List: $24.00 MAA Member: $18.50

Two of the most fascinating problems to challenge mathematicians in recent years concern the construction of data transmission codes that can correct errors introduced by static and the search for efficient ways to pack ping-pong balls into a box. Can one design the best error-correcting codes? Can one find the most efficient sphere packing?

Therein lies a fascinating story which is told with great skill and clarity in this important addition to the Carus Mathematical Monograph series. The author has packed (sic) into 175 pages all of the basic mathematical ideas of this saga woven into a gripping historical account of the journey from error-correcting codes to sphere packings to simple groups.

**Table of Contents**





# MAA STUDIES IN MATHEMATICS

## Studies in Numerical Analysis

*MAA Studies in Mathematics #24*

Gene H. Golub, Editor
415 pp. Hardbound.
List: $42.00  MAA Member: $31.00

This volume is a collection of papers describing the wide range of research activity in numerical analysis. The articles describe solutions to a variety of problems using many different kinds of computational tools. Some of the computations require nothing more than a hand held calculator: others require the most modern computer. While the papers do not cover all of the problems that arise in numerical analysis, they do offer an enticing and informative sample.

Numerical analysis has a long tradition within mathematics and science, beginning with the work of the early astronomers who needed numerical procedures to help them solve complex problems. The subject has grown and developed many branches, but it has not become compartmentalized. Solving problems using numerical techniques often requires an understanding of several of the branches. This fact is reflected in the papers in this collection.

Computational devices have expanded tremendously over the years, and the papers in this volume present the different techniques needed for and made possible by several of these computational devices.

**Table of Contents**